ABSTRACT

Interval-Censored Negative Binomial Models: A Bayesian Approach

Stephanie M. Doherty, Ph.D.

Chairperson: John W. Seaman, Jr.

Count data are quite common in many research areas. Interval-censored counts, in which an interval representing a range of counts is observed rather than the precise count, may arise in many situations, including survey data. In this dissertation we develop a model for accommodating interval-censored count data through the interval-censored negative binomial model, with an expansion to a regression model in which the interval count responses are regressed on covariate values. We employ both frequentist and Bayesian methods to arrive at point and interval estimates for the negative binomial parameters. We find that many factors, including the interval-censored widths and the tendency of the precise counts toward either endpoint of the intervals, affect parameter estimates based on interval-censored data as compared to estimates using only precise data. We perform simulation studies in the non-regression and regression contexts, which compare the interval-censored model to alternatives for accommodating interval-censored data. These methods are precise-count analyses based on the lower endpoints, upper endpoints, or means of the observed intervals. For the scenarios in our simulation experiments, we find that the interval-censored model outperforms the lower endpoint and upper endpoint methods, and performs at least as well as, or better than, the mean method. We conclude with an extended example, in which we compare the interval-censored method to the lower and upper endpoint methods for health-related quality of life survey data that are interval-censored. We find that the interval-censored method allows us to calculate parameter estimates and conduct posterior inferences, without the need to discard any information provided in the study.

Interval-Censored Negative Binomial Models: A Bayesian Approach

by

Stephanie M. Doherty, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

Jack D. 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

Jane L. Harvill, Ph.D.

David J. Kahle, Ph.D.

James D. Stamey, Ph.D.

Janet R. Crow, Ph.D.

Accepted by the Graduate School
August 2012

J. Larry Lyon, Ph.D., Dean

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I am deeply grateful to my parents. You have provided me with so many opportunities in this life, and I would not be where I am today without the outpouring of love and support that you have shown. It means a lot just to know that you are proud of me; you have expressed your pride in my accomplishments and your unconditional love throughout my life.

Thank you to my advisor, Dr. John Seaman. You have provided wisdom and guidance along every step of this journey, from all the details of the research to the interviewing and job selection process. You have always given positive feedback, helping me to believe in my abilities as a statistician. I could not have reached this point without your abundant assistance.

To my husband, Brad, thank you for your patience and loving support as I have worked toward my goal, through my long hours of work and occasional frustrations. I cannot wait for the next phase in our life and all of the adventures that we will take on together. Thank you as well to all of my friends and family, who have helped shape me into who I am today.

Thank you to the Baylor University Department of Statistics faculty and fellow students. I have learned so much in these seven years, from my time as an undergraduate to these final days in the program. You have helped me to grow as a statistician, as a lifetime student, and as an individual. Additionally, I would like to thank Mike Hutcheson in the Baylor Academic and Research Computing Services department for working many hours to resolve my computer system issues, which allowed me to complete my simulations in a timely manner.

Above all, to God be the glory, for all I have and all I am are because of His infinite grace.

DEDICATION

To Brad

my best friend

my love

CHAPTER ONE

Introduction

Interval-censored data may arise in a variety of applications, including quite commonly in the context of survival and reliability analyses. In a typical clinical trial with time-to-event data, there are a fixed number of patients at the start to which a treatment is applied. Due to time or cost constraints the investigator may terminate the study before all patients realize the event of interest. Those subjects who have yet to experience the event at trial's end are considered right-censored, with recorded times equal to the duration of the study period. On the other hand, if an individual in a study has already experienced the event of interest before that person is observed in the study, then that individual's time-to-event is unknown and considered left-censored. The third type, interval-censoring, is more general and occurs when patients have periodic follow-up and a given patient's event time is only known to fall in an interval $(L_i, R_i]$. Here L_i and R_i denote the left and right endpoints of the censoring interval (Klein and Moeschberger 2003).

Though censoring is often encountered in the survival context, it also exists in other types of data, including discrete counts. Pruszynski (2010) developed the interval-censored binomial model in her dissertation, using frequentist and Bayesian methods to address this problem. She also introduced likelihood models for a single interval-censored Poisson count and a single interval-censored negative binomial count. Watson (2011) expanded on the Poisson case, considering multiple observations and introducing several Poisson regression models for interval-censored data. In this dissertation we expand on the interval-censored negative binomial case. We give some motivating examples for using this type of model in Sections 1.1 and 1.2.

1.1 General Social Survey Example

Ntzoufras (2009) analyzes data from $n = 550$ respondents to the 1990 U.S. General Social Survey (GSS). The GSS is conducted by the National Opinion Research Center (NORC) and began in 1972 for the purpose of "monitoring societal change and the growing complexity of American society" (NORC 2012). In one particular question on the 1990 survey, respondents were asked to report their total number of sexual intercourses in the previous month. Such information is useful, for example, in epidemiological studies of sexually transmitted diseases. Ntzoufras regressed these precise-count answers on the respondents' genders in a negative binomial generalized linear model (GLM). Though precise counts were given in this data set, we might imagine a scenario in which the researcher does not require the respondent to give a precise answer on data of this nature. Then each respondent may give a range, i.e., "I know it was between 3 and 5," and the researcher would be presented with an interval-censored count data set. Then using our interval-censored negative binomial model, we could perform a regression analysis—in this case, with gender as the single binary covariate. In Chapter 3 we present Ntzoufras's original analysis and then generate hypothetical interval-censored data. There we compare regression analyses on the precise count responses as well as the interval-censored count responses, using both frequentist and Bayesian methods.

1.2 The Health-Related Quality of Life Survey

Consider the following scenario, in which the interval-censored negative binomial model would be useful: The Centers for Disease Control and Prevention (CDC) have developed several survey questions that attempt to measure respondents' health-related quality of life (HRQOL). Quality of life (QOL) is a term that "conveys an overall sense of well-being, including aspects of happiness and satisfaction with life as a whole" (CDC 2000). Then health-related quality of life refers to those aspects of QOL that have a direct impact on health—physical or mental. Measuring and tracking HRQOL has become an important goal as several national agencies have incorporated HRQOL improvement into their health objectives.

As part of HRQOL assessment, the CDC developed a concise set of questions called the Healthy Days measures, which state and community healthcare organizations could use as a standard for measuring each respondent's perceived health status over time. These four questions assess the individual's 1) self-rated health, 2) number of days in the recent past of physical unwellness, 3) number of recent days of mental unwellness, and 4) number of recent days when activity was limited due to poor physical or mental health. The second and third questions read as follows (CDC 2000):

> Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?
>
> Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

From these questions, the CDC creates a summary index of unhealthy days. This summary index is an estimate of the overall number of days in the previous 30 days when the patient perceived either physical or mental unwellness. To calculate this comprehensive score, the CDC uses the maximum of 30 days and the sum of the responses to the two questions above. For example, a person who reports 4 physically unhealthy days and 2 mentally unhealthy days would be assigned a summary score of 6. A person reporting 20 physically unhealthy days and 15 mentally unhealthy days will receive a score of 30.

This summary index assumes no overlap in reported physically and mentally unhealthy days, aside from the cases where the sum exceeds 30, for which minimal overlap is assumed. However, we might imagine that some respondents experienced mental unwellness coinciding with their days of physical unwellness. If a person reports 15 days for each question, the CDC will give a summary score of 30 unhealthy days for this person, but at the other extreme this person may have experienced only 15 unhealthy days. There also may have been some other degree of overlap, which would place the number of unhealthy days somewhere between 15 and 30. In any case, it should be preferable to acknowledge the inherent uncertainty in this score when using such data. The interval-censored negative binomial model can help us accomodate this uncertainty. Our application in Chapter 4 will make use of responses to this question as given on the 2003 California Health Interview Survey (CHIS).

1.3 Plan for the Dissertation

The dissertation is organized into three main chapters. In Chapter 2 we introduce the interval-censored negative binomial likelihood. We demonstrate the behavior of the likelihood contour surface with respect to several aspects of interval-censored modeling:

- The interval-censored widths
- The percentage of data that are interval-censored within a given set
- The sample size of the data set
- The positioning of the intervals relative to the true precise counts.

Each of these factors plays a role in the behavior of the interval-censored likelihood surface relative to a corresponding precise likelihood surface. In Chapter 2 we develop maximum likelihood estimation for the model using the Newton-Raphson iterative method. Estimates and standard errors from the Newton-Raphson method are used to obtain Wald interval estimates for the parameters. However, the Wald interval is not always appropriate for certain parameters based on the sampling distributions of the MLEs, and we explore this fact for the interval-censored negative binomial model. We then introduce a Bayesian model for analysis, including discussion of several prior distribution choices. We apply a Bayesian analysis to several of the data examples previously presented, and we conclude the chapter with a simulation study.

In Chapter 3 we move into the regression context for the interval-censored negative binomial model. We define the regression model specifically for the GSS example in Section 1.1, then analyze this precise-count data via maximum likelihood using the Newton-Raphson method, and Bayesian estimates using Markov chain Monte Carlo (MCMC). We then generate a smaller precise sample based on the GSS data parameter estimates, and create interval-censored data around this set. We compare frequentist and Bayesian estimates for precise counts with mild and intermediate interval censoring, then re-analyze in the Bayesian framework for several choices of prior distributions. We conclude with a discussion of the likelihood behavior for several data scenarios as well as a simulation study of model operating characteristics.

Chapter 4 is an applied chapter, in which we apply our frequentist and Bayesian techniques to a subset of real data from the CHIS. We compare frequentist and Bayesian estimates based on three possible data sets for the same survey responses: two precise data sets using either extreme of totaling unhealthy days based on mentally and physically unhealthy days, and an interval-censored approach which allows the intervals to range between these extremes.

CHAPTER TWO

The Interval-Censored Negative Binomial Model

Count data are quite common in many areas of research, and interval-censored counts arise, as we saw in Chapter 1, for certain types of survey data. The Poisson model is commonly used in a preliminary approach to count data. Watson (2011) introduces and develops a model for Poisson-distributed data that are interval-censored. The primary limitation in using a typical Poisson model to fit count data is the required assumption that the variance of the counts, $V(Y)$, is equal to the mean of the counts, $E(Y)$. Count data that do not satisfy this equidispersion assumption are common. The negative binomial distribution is one means of accomodating this type of data. In this dissertation we consider count data that are over-dispersed, i.e. $V(Y) > E(Y)$, with at least one of the counts in a given data set being interval-censored. Our proposed model for handling this type of data is the interval-censored negative binomial model, which we introduce in Section 2.2. Our development of this model begins with a review of the negative binomial model, in Section 2.1.

2.1 The Negative Binomial Distribution

The negative binomial model may be more appropriate than a Poisson model in certain scenarios, especially when the variance of the data exceeds the mean, as the Poisson model assumes that these two quantities are equal. Overdispersion (or underdispersion) is accomodated through the use of a second parameter in the negative binomial model, called the dispersion parameter.

Ntzoufras (2009, p. 283) defines a discrete random variable Y as having a negative binomial distribution, $Y \sim NB(\pi, r)$, if its mass function is given by

$$f(y;\pi,r)=\frac{\Gamma(y+r)}{y!\Gamma(r)}\pi^r(1-\pi)^y, \tag{2.1}$$

for $y=0,1,2,\dots,r$, with $r>0$ and $0\leq\pi\leq 1$. Then the mean and variance are $r(1-\pi)/\pi$ and $r(1-\pi)/\pi^2$, respectively.

The parameter r need not be an integer. However, a common use of the negative binomial distribution is modeling the number of times needed to repeat a Bernoulli experiment with success probability π until a specified number of successes is reached. If we call this target number N, then $r=N$ will be an integer and $N+y$ is the total number of times needed to repeat the Bernoulli experiment. Ntzoufras (2009) notes that in the more general case, r is a positive real number and the negative binomial is used to model overdispersed count data. The ratio of the variance to the mean, the *dispersion index*, is given by

$$DI\equiv\frac{\mathrm{Var}(Y)}{E(Y)}=\frac{1}{\pi}.$$

In this more general case it may also be advantageous to use the parameterization $\lambda=r(1-\pi)/\pi$, so that $E(Y)=\lambda$ and $V(Y)=\lambda+\lambda^2/r$, with $DI=1+\lambda/r$. This is the parameterization that results from deriving the negative binomial model via the Poisson-gamma mixture model

$$Y|U=u\sim\text{Poisson}(\lambda u)\text{ and }U\sim\text{gamma}(r,r),$$

where gamma(r,r) denotes the gamma density given by

$$\frac{r^r}{\Gamma(r)}u^{r-1}\exp(-ru).$$

The marginal distribution of Y that results is given by

$$\begin{aligned}f(y)&=\int_0^\infty f(y|u)f(u)du\\&=\frac{\Gamma(y+r)}{y!\Gamma(r)}\left(\frac{r}{r+\lambda}\right)^r\left(\frac{\lambda}{r+\lambda}\right)^y,\end{aligned}$$

which is a negative binomial distribution with parameters $\pi=r/(r+\lambda)$ and r.

Under this parameterization, note the inverse relationship of the variance with the dispersion parameter, $V(Y) = \lambda + \lambda^2/r$. As the dispersion parameter r increases, the variance decreases—and approaches the variance of a Poisson distribution. Hilbe (2011) notes that an alternative parameterization using α as the dispersion parameter has been recently preferred, including in the major software implementations of the negative binomial. This mass function is given as

$$f(y;\lambda,\alpha) = \frac{\Gamma(y+1/\alpha)}{y!\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1+\alpha\lambda}\right)^{y}. \tag{2.2}$$

With this parameterization the mean is again λ, and the variance is $\lambda + \alpha\lambda^2$. Now we see a direct relationship between the dispersion parameter, α, and the variance of Y. This form is especially preferred, according to Hilbe (2011), when modeling overdispersed Poisson count data. In this form, α is directly related to the amount of overdispersion in the data. If there is no overdispersion in the data, i.e. the data are Poisson-distributed, then α will equal zero. Increasing α increases overdispersion. In practice, values for α typically range from 0.01 to around 4. We employ this parameterization throughout the dissertation, referring to the distribution as NB(λ, α).

There is a complex interplay between λ, α, $V(Y)$, and the shape of the respective likelihoods. We have that $\lambda = (1-\pi)/\alpha\pi$, so that α decreases as λ increases. We also have $E(Y) = \lambda$ and $V(Y) = \lambda + \alpha\lambda^2$. Thus, λ and α are directly related to the variance of Y, and this in turn will have an impact on their joint likelihood. Increasing either α or λ will increase the variance of Y, and yet increasing the value of one of these parameters will decrease the value of the other. To visualize the interplay, we plot $V(Y)$ as a function of both α and λ in Figure 2.1. We start by fixing the negative binomial probability, π, to be 0.1. Letting α range from 0 to 4 in our plot constrains λ for fixed π because $\lambda = (1-\pi)/\alpha\pi$. The highest black curve represents the variance $V(Y)$ versus the dispersion parameter α and the mean λ for

$\pi = 0.1$. The red curve just below is the same relation for $\pi = 0.2$. This continues for $\pi = 0.3, \ldots, 1$. In this plot we see that λ has a large impact on the variance, although one should note the difference in scale between λ and α. The variance in this representation actually decreases for large α, because a large α is overshadowed by a correspondingly smaller λ, which leads to smaller variance. Later, we shall see that these relationships have considerable impact on the performance of our estimators.

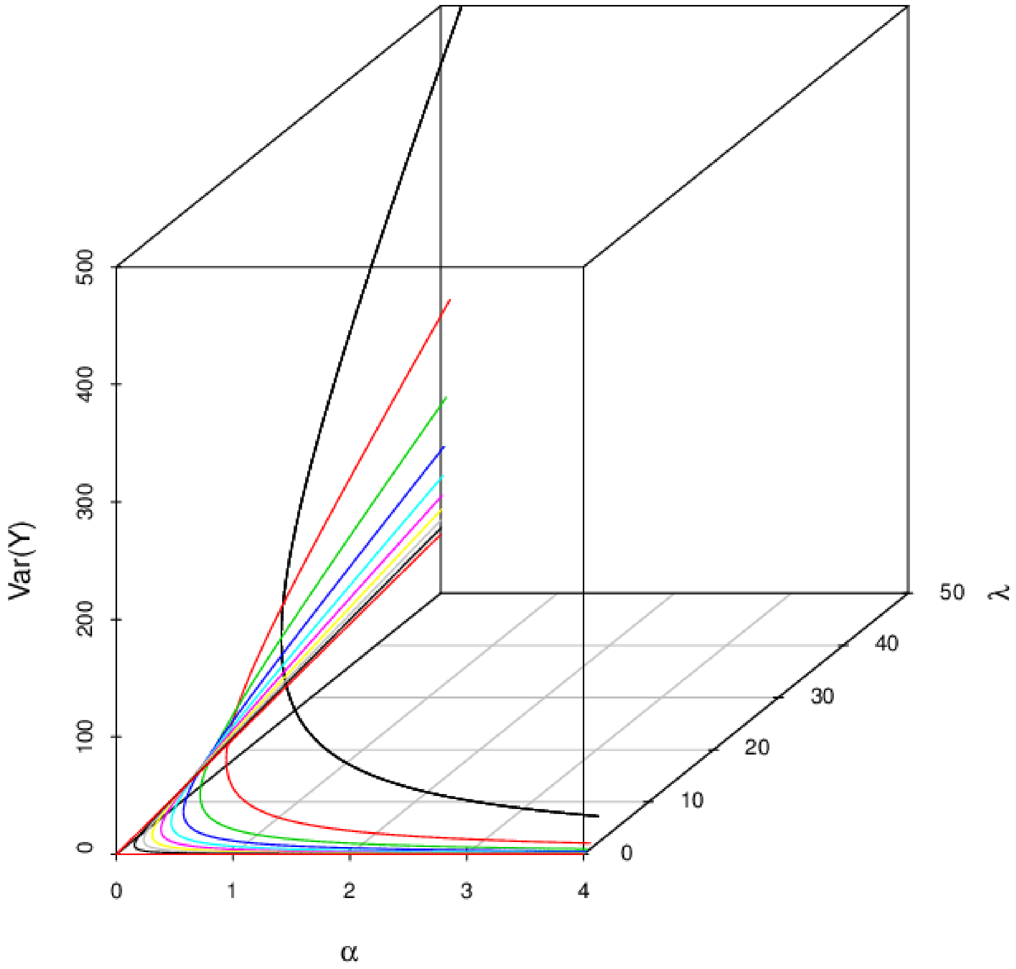

Figure 2.1: Variance as a function of α and λ of a negative binomial random variable Y. The probability π is fixed along each curve, starting at the highest black curve where $\pi = 0.1$, and increasing π by 0.1 until $\pi = 1$, yielding the lowest red curve corresponding to the curve.

2.2 The Interval-Censored Negative Binomial Model

We will utilize the parameterization in (2.2) for our interval-censored model. Suppose now that we no longer know the precise count Y but we assume that Y falls in the interval $[j, k]$ with probability 1, where j and k are positive integers and $Y = j, j+1, \ldots, k$. We refer to $[j, k]$ as the censoring interval. We refer to $k - j$ as the censoring interval width. Then the censored likelihood function for a single observation, modified from Pruszynski (2010) to fit our parameterization, is

$$L(\lambda, \alpha | j \leq y \leq k) = \sum_{y=j}^{k} \frac{\Gamma(y + 1/\alpha)}{y!\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1+\alpha\lambda}\right)^{y}. \tag{2.3}$$

Now suppose there are n observations $[j_i, k_i]$, $i = 1, \ldots, n$, where $j_i \leq k_i \; \forall \; i$. Denote by $\mathbf{d}_n$ the collection of these intervals. That is, $\mathbf{d}_n = \{[j_i, k_i]\}_{i=1}^{n}$. The likelihood for λ and α given $\mathbf{d}_n$ is

$$L(\lambda, \alpha | \mathbf{d}_n) = \prod_{i=1}^{n} \sum_{y_i=j_i}^{k_i} \frac{\Gamma(y_i + 1/\alpha)}{y_i!\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1+\alpha\lambda}\right)^{y_i}. \tag{2.4}$$

Intuitively we expect that an interval-censored likelihood will display more dispersion than a corresponding likelihood in which all counts are precisely known. However, this is not always the case. The shape of an interval-censored likelihood depends on a variety of factors including the true values of the parameters from which the data were generated, the sample size, the width of the intervals in the interval-censored data, the positioning of the true values within the intervals, and the percentage of data that are interval-censored. We explore these concepts with some examples in the following section, with an eye toward identifying parameter values for design points in later simulation studies.

2.3 Likelihood Examples

Data that are interval-censored contain more uncertainty and thus should have less informative likelihoods than a corresponding set of precise data values. However, in Figure 2.2 we illustrate an example in which the precise and interval-censored likelihood surfaces are virtually identical. We created this set of likelihood contours by generating $n = 10$ exact counts from a $NB(10, 0.1)$ distribution to form y. We then created an interval-censored data set by setting a common interval censoring width to 1 and randomly selecting whether $y_i = j_i$ or $y_i = k_i$ for each set. For example, the first precise point was $y_1 = 9$ and we generated $[j_1, k_1] = [9, 10]$; the second precise point was $y_2 = 19$ and we generated $[j_2, k_2] = [18, 19]$, etc.

Because the negative binomial distribution is affected by the values of both α and λ, we must consider likelihood contours in our examples. The curves plotted on the margins are the profile likelihoods for λ and α separately. In general, if (θ, η) is the full parameter vector, and θ is the parameter of interest, then given the joint likelihood $L(\theta, \eta)$, the profile likelihood of θ is (Pawitan 2001, pp. 61-62)

$$L(\theta) = \max_{\eta} L(\theta, \eta).$$

Thus to find the profile likelihood for λ, we find the MLE, $\hat{\alpha}$, for α and plot $L(\lambda, \hat{\alpha})$, which gives the likelihood curve across the top margin of the contour plot. Likewise, to find the profile likelihood for α we find the MLE for λ, $\hat{\lambda}$, and plot $L(\hat{\lambda}, \alpha)$ across the right margin of the contour plot.

This example demonstrates that, even though our interval-censored data set contains less information than the precise set, we lost little precision and created little "bias," at least with narrow censoring intervals. Here, by "bias," we mean the positioning of the likelihood contours relative to the true parameter values used to generate the data.

As we might expect, increasing the widths in the interval-censored data leads to more differentiation between the precise and interval-censored likelihoods. Suppose y_i is an interval censored observation such that $y_i \in [j_i, k_i]$ with probability one. Denote the interval censoring width of the ith observation by w_i; that is, $w_i = k_i - j_i$. In Figure 2.3, the three panels exhibit the difference between the precise likelihood ($w_i = 0 \ \forall \ i$) and each of three common interval censoring widths: $w_i = 1, 5$, and

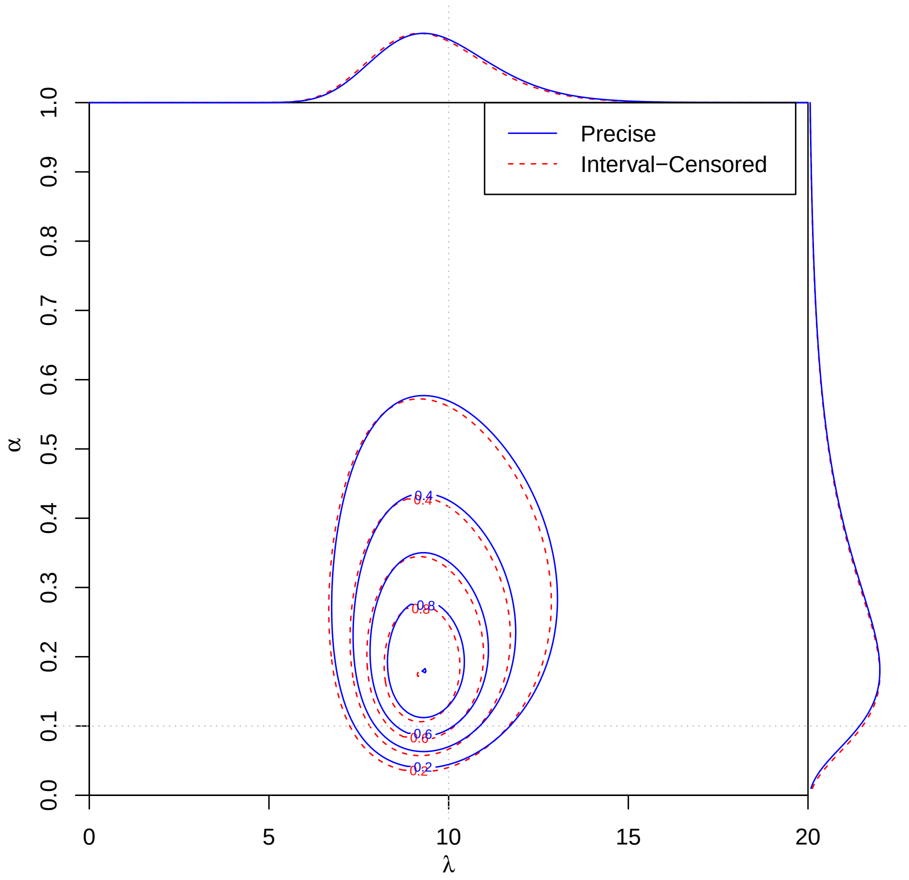

Figure 2.2: Likelihood contours for interval-censored (dashed red curves) and precise (solid blue curves) data, $n = 10$, $\alpha = 0.1$, $\lambda = 10$. The profile likelihood of λ is given on the upper axis of the graph. The profile likelihood of α is provided outside the right axis of the graph.

10. The latter case represents data that would be unlikely in practice, as the widths match the expected count λ, but we plot it here for illustration purposes to show the effect of increasing the widths. Notice that as the widths increase, the interval-censored data produces more uncertainty in the profile likelihoods for both λ and α.

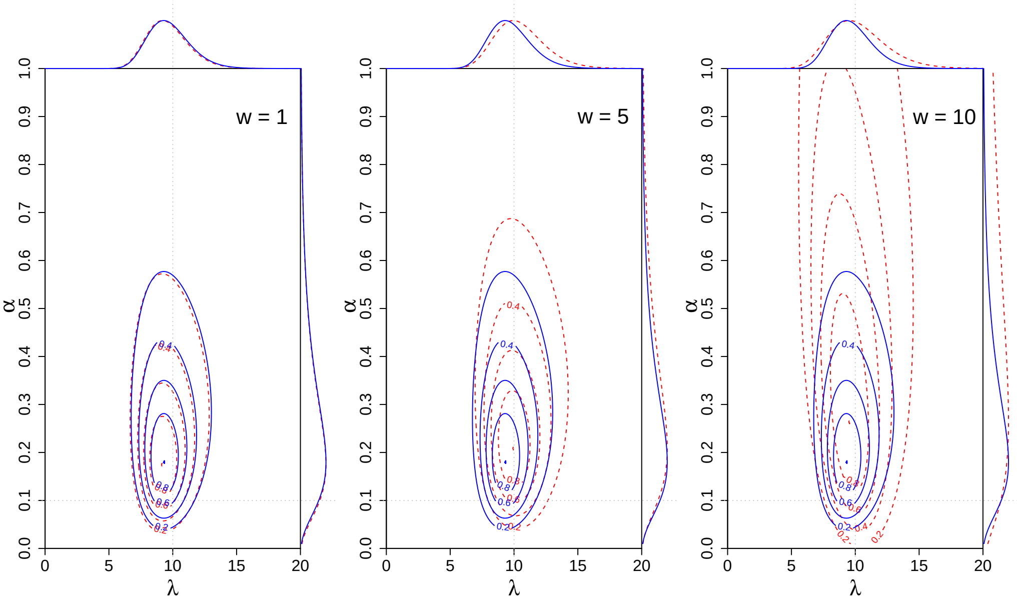

Figure 2.3: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, $n = 10$, $\alpha = 0.1$, $\lambda = 10$. Profile likelihoods of λ on upper external axes and profile likelihoods of α on outer right axes. The left panel has $w = 1$; center panel $w = 5$; and right panel $w = 10$.

The percentage of interval-censored observations in a data set will clearly affect the likelihood. For the example in Figure 2.4 we examine the effect for 5%, 50%, 75% and 100% interval-censored data. The precise data are a sample of size 20 generated from a $NB(10, 1)$ distribution. In each interval-censored set, the points that are not interval-censored are equal to the corresponding precise points. Increasing the percentage of interval-censoring changes both the location and the scale of the likelihood contours.

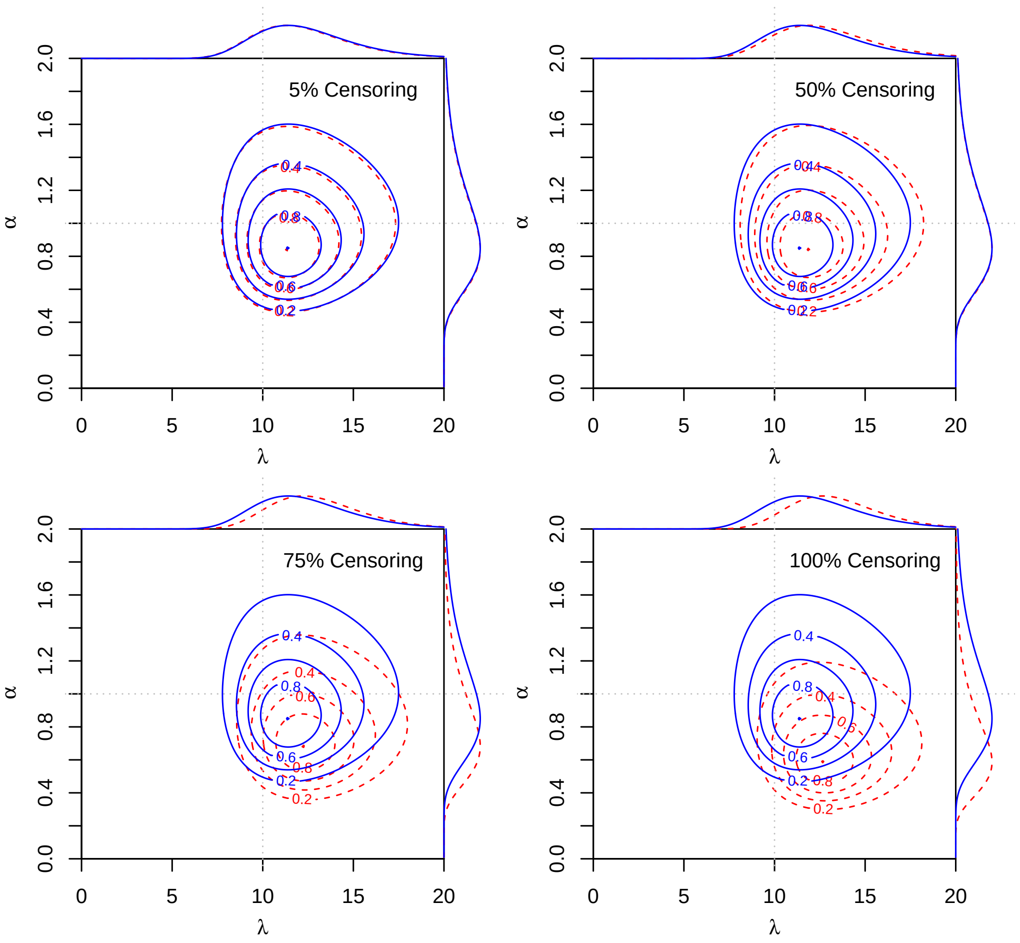

Figure 2.4: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, $\alpha = 1$, $\lambda = 10$, interval-censored widths equal 5. Percentage of interval censoring is 5% in the top left panel, 50% in the top right panel, 75% in the bottom left panel, and 100% in the bottom right panel. Profile likelihoods for λ are on the external top axes; profile likelihoods for α are on the external right axes.

Sample size has an effect on the location and scale of the likelihood, as we might expect, and also affects the differentiation between the interval-censored and precise cases. In Figure 2.5 we first generate $n = 5$ observations from a $NB(10, 2)$. The interval-censored data set is created with widths $w_i = 5$ for all i, and the position of each y_i within $[j_i, k_i]$ randomly generated. The process is repeated in the center plot for $n = 20$, then in the right plot for $n = 50$.

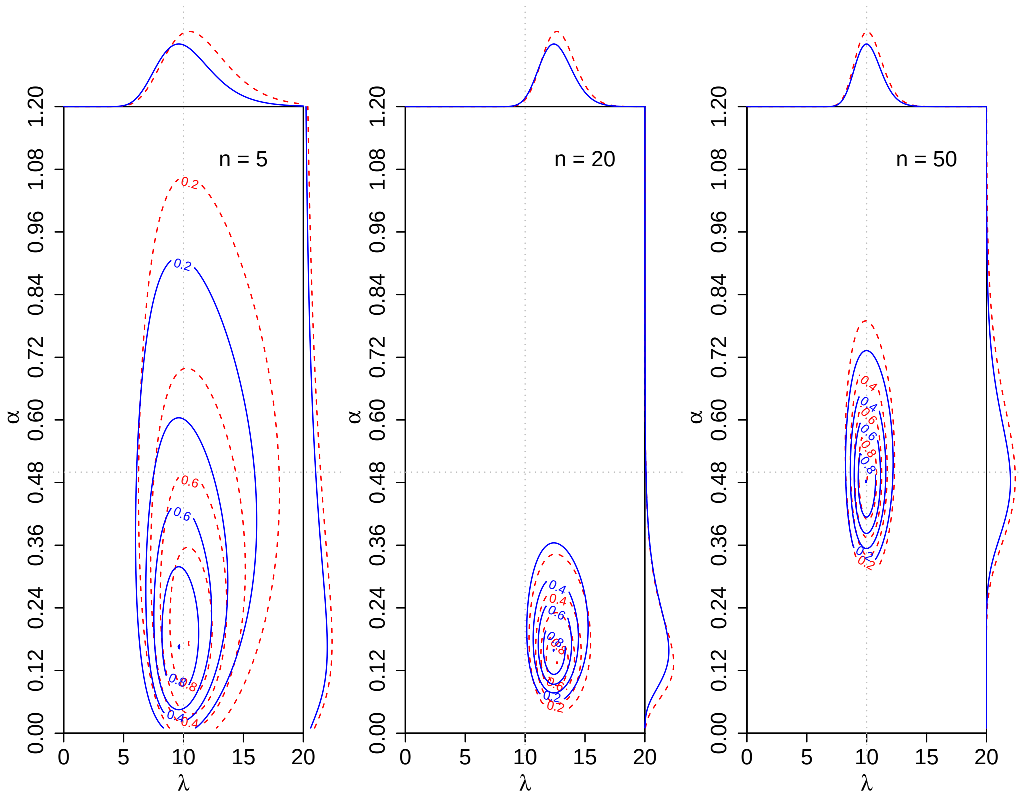

Figure 2.5: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, $\alpha = 2$, $\lambda = 10$, interval-censored widths equal 5. Sample size n is 5 (left panel), 20 (center panel), and 50 (right panel). The profile likelihoods for λ are on the external top axes of the plots; for α the profile likelihoods are on the right external axes of the plots.

As the sample size increases, the likelihood contours under in the precise and interval-censored cases shift closer to the underlying true values from which we generated the data, and the contours decrease in dispersion. We see differentiation between the precise and interval-censored models for $n = 5$, but for $n = 50$ the fact that the interval-censored values are 5 units wide does not seem to have an effect on the overall shape—the interval-censored likelihood has small bias relative to the precise likelihood and similar variance.

We must also consider the positioning of the precise value, y_i, within the interval $[j_i, k_i]$ as this has an effect on the shape of the likelihood. In Figure 2.6 we artificially assume we know the actual count and observe what happens when the intervals fall in different positions relative to the precise observation. We generate the precise data, $n = 20$, from a $NB(10, 0.5)$ distribution. In the first plot we set $j_i = y_i$ and set $w_i = 6$ so that $k_i = y_i + 6 \ \forall \ i$. In the second plot, we center the interval over the precise value, so that $j_i = y_i - 3$ and $k_i = y_i + 3 \ \forall \ i$. In the last plot, we set $k_i = y_i \ \forall \ i$ and let $j_i = y_i - 6 \ \forall \ i$. If this results in $j_i < 0$, we set $j_i = 0$ and let $k_i = 6$ to maintain constant width across observations.

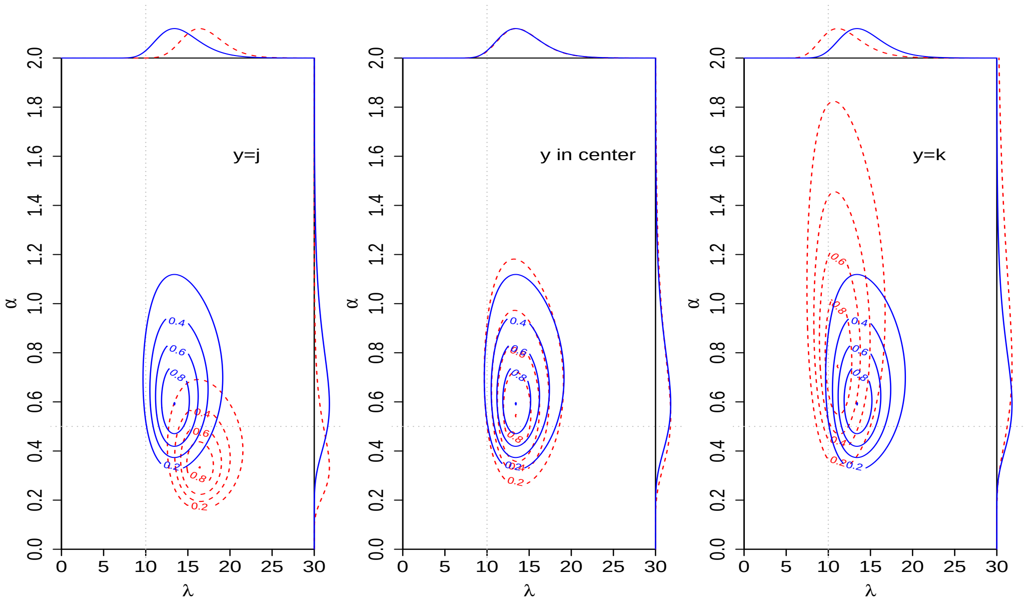

Figure 2.6: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, $\alpha = 0.5$, $\lambda = 10$, interval-censored widths equal 6. Interval tendency changes across the three plots, from $y = j$ (left panel) to y centered between j and k (center panel), to $y = k$ (right panel). The profile likelihoods for λ are on the outer top axes of the plots and for α are on the outer right axes of the plots.

When y_i tends to the lower limit of the intervals, corresponding to survey respondents giving an over–estimate with their intervals, the likelihood is peaked over larger values of λ and smaller values of α relative to the precise likelihood. Decreased α also leads to decreased variance, so the interval-censored likelihood is less dispersed than in the precise case, and further from the true generating values of α and λ. Thus, there is a tradeoff in this particular example: increased precision for additional bias. When j and k are such that y is centered in the interval, as in the second plot, we see that the profile likelihood for λ is similar for the interval-censored and precise cases. The profile likelihood for α is positively skewed because there is more dispersion in the interval-censored case. Finally, when y falls on the upper limits of the intervals, corresponding to respondents giving an under-estimate with their intervals, the profile likelihood for α shifts to the right and the profile likelihood for λ shifts to the left. The profile likelihood for λ appears to have as much or less variability than in the precise case. This seems most likely related to the decrease in λ and consequent smaller variance. Variance increases as we increase α.

2.4 Maximum Likelihood Estimation

In this section we consider maximum likelihood estimation of the interval-censored negative binomial regression model. There are iteratively re-weighted least squares (IRLS) methods available for the canonical and traditional parameterizations of the negative binomial model, but Hilbe (2011, p. 207) recommends against their use in specialized parameterizations like the one we employ. Instead, we use a Newton-Raphson method to obtain maximum likelihood estimates of λ and α in our interval-censored model. We also consider Wald approximations of the standard errors of these estimators.

2.4.1 Newton-Raphson and Wald Approximations

Suppose a model depends on a $p \times 1$ vector of parameters, $\boldsymbol{\beta}_r$. Let $\boldsymbol{\mathcal{L}}$ be the log-likelihood for $\boldsymbol{\beta}_r$. The Newton-Raphson method uses the observed information matrix, $\mathbf{H}$, and the gradient vector $\mathbf{U}$ of the log-likelihood, which is the vector of first partial derivatives of the log-likelihood. The Newton-Raphson algorithm calculates the rth iteration of parameter estimates, $\boldsymbol{\beta}_r$, by

$$\boldsymbol{\beta}_r = \boldsymbol{\beta}_{r-1} - \mathbf{H}^{-1}\mathbf{U} \tag{2.5}$$

where $\mathbf{H} = \mathbf{H}_{r-1}$ and $\mathbf{U} = \mathbf{U}_{r-1}$. The Newton-Raphson method begins with an initial estimate, $\boldsymbol{\beta}_0$, and iteration continues until a pre-specified threshold on the absolute difference $|\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r-1}|$ is attained.

The negative binomial log-likelihood for λ and α given precise data $\mathbf{y} = (y_1, \dots, y_n)$ is

$$\mathcal{L}(\lambda, \alpha|\mathbf{y}) = \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right) - \frac{1}{\alpha}\ln(1+\alpha\lambda) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right). \tag{2.6}$$

Then

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \sum_{i=1}^{n} \frac{y_i - \lambda}{\lambda(1+\alpha\lambda)}, \tag{2.7}$$

and

$$\frac{\partial\mathcal{L}}{\partial\alpha} = \sum_{i=1}^{n} \left[\frac{1}{\alpha^2}\left(\ln(1+\alpha\lambda) + \frac{\alpha(y_i-\lambda)}{1+\alpha\lambda}\right) + \Psi\left(y_i + \frac{1}{\alpha}\right) - \Psi\left(\frac{1}{\alpha}\right)\right], \tag{2.8}$$

where Ψ is the the digamma function—the derivative of the log-gamma function.

The MLEs $\hat{\lambda}$ and $\hat{\alpha}$ for a given iteration are found by setting (2.7) and (2.8) equal to zero and solving for λ and α, respectively. Then the Hessian matrix $\mathbf{H}$ for the Newton-Raphson method is the matrix of observed second partial derivatives,

$$\begin{bmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial \hat{\lambda}^2} & \dfrac{\partial^2 \mathcal{L}}{\partial \hat{\alpha} \partial \hat{\lambda}} \\ \\ \dfrac{\partial^2 \mathcal{L}}{\partial \hat{\lambda} \partial \hat{\alpha}} & \dfrac{\partial^2 \mathcal{L}}{\partial \hat{\alpha}^2} \end{bmatrix}.$$

The Newton-Raphson criteria for convergence are a specified maximum difference in log-likelihood functions as well as a maximum difference in parameter estimates. Initial values may be chosen in several ways: setting all initial values to zeros or ones, or using estimates from a Poisson model for the same data. The algorithm is as follows (Hilbe 2011, p. 56):

(1) Initialize parameters $\boldsymbol{\beta}_0$.

(2) Set $\boldsymbol{\beta}_{n-1} = \boldsymbol{\beta}_0$.

(3) Calculate $\mathbf{U}$ from $\boldsymbol{\beta}_{n-1}$.

(4) Calculate $\mathbf{H}$ from $\boldsymbol{\beta}_{n-1}$.

(5) Calculate $\boldsymbol{\beta}_n = \boldsymbol{\beta}_{n-1} - \mathbf{H}^{-1}\mathbf{U}$.

(6) Calculate new log-likelihood $\mathcal{L}_n = \mathcal{L}_{n-1}$.

(7) Check to see if the elementwise differences abs$(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n-1})$ and abs$(\mathcal{L}_n - \mathcal{L}_{n-1})$ are within the specified tolerance levels. If not, repeat steps 3-7 with $\boldsymbol{\beta}_n$ from step 5 as the new $\boldsymbol{\beta}_{n-1}$.

The gradient terms for $\mathbf{U}$ in the precise case are as given in (2.7) and (2.8), and the closed-form second derivatives for $\mathbf{H}$ are also available in the precise case (see, for example, Hilbe 2011 pp. 192-193). These expressions for the interval-censored log-likelihood are not available in closed form. The *R* program for Newton-Raphson maximization, `maxNR`, accomodates this by estimating the finite-difference gradient and Hessian at each iteration.

Once the differences in parameter estimates and log-likelihood functions are within their respective tolerances, the algorithm stops, yielding $\hat{\lambda}$ and $\hat{\alpha}$. The final

Hessian matrix calculated is used to find the standard errors for $\hat{\lambda}$ and $\hat{\alpha}$. The approximate standard errors are equal to the square roots of the diagonal terms of the negative inverse Hessian matrix.

Consider the first set of data plotted in Figure 2.3, in which $w_i = 1 \; \forall \; i$. Based on the estimates and standard errors we obtain from the Newton-Raphson technique, the 95% Wald confidence intervals under the precise and interval-censored scenarios are given in the second and third columns of Table 2.1. In the precise and interval-censored cases, lower bounds for α were calculated to be -0.07, but have been set to zero since $\alpha > 0$.

Table 2.1. 95% Wald Interval Estimates for Precise and Interval-Censored Data

Param	Truth	Precise CI	IC $w_i = 1 \; \forall \; i$	IC $w_i = 10 \; \forall \; i$
λ	10	$(6.18, 12.42)$	$(6.14, 12.25)$	$(5.21, 13.84)$
α	0.1	$(0.00, 0.42)$	$(0.00, 0.42)$	$(0.00, 0.84)$

The precise and interval-censored Wald interval estimates for this set of data are very similar, as expected given the similarity in shape of the likelihood contours. Now we apply the same estimating algorithm to the data in the third contour plot of Figure 2.3. Here the precise data, $\mathbf{y}$, remained the same, while we increased the interval-censored widths to 10 which, as already noted, is an extreme case. The 95% interval-censored interval estimates are given in the fourth column of Table 2.1. Here the actual calculated lower bound for α was -0.30, and so was set to zero. As we would expect based on the contours in the third plot in Figure 2.3, the widths of the intervals on both λ and α have increased.

Wald interval estimates are based on large-sample approximation theory; in Section 2.4.2 we demonstrate the appropriateness of using Wald interval estimates, as calculated in Table 2.1, for the small-sample data represented in Figure 2.3. Rather than serving as a full exploration of when the approximation might be appropriate, this section demonstrates situations where we might choose to use a Wald interval, and situations where it would not be ideal, within a single data scenario.

2.4.2 Appropriateness of Wald Interval Estimates

For an estimator, $\hat{\theta}$, Wald intervals are based on the assumption that, as the sample size becomes large, the sampling distribution of $\hat{\theta}$ may be well approximated by

$$\hat{\theta} \sim N(\theta, I(\hat{\theta})^{-1}), \tag{2.9}$$

where $I(\hat{\theta})$ is the observed information; see, for example, Pawitan (2001), pp. 247-250. If we believe the normal approximation for the sampling distribution of $\hat{\theta}$ is suspect, then the standard Wald formula may be a poor choice for the approximate confidence interval, as it is based on the quadratic approximation providing a good fit to the likelihood (Pawitan, 2001, p. 241). Consider again the data from Figure 2.3. In order to illustrate the suitability of the normal approximation to the sampling distributions of $\hat{\lambda}$ and $\hat{\alpha}$, we generated 300 datasets of size 10 from a $NB(10, 0.1)$ distribution. For each precise data set we also generated interval-censored data with $w_i = 1 \; \forall \; i$, and then with $w_i = 10 \; \forall \; i$. We find $\hat{\lambda}$ and $\hat{\alpha}$ for the precise and interval-censored data in each replication based on the Newton-Raphson iterative scheme, then plot histograms and normal quantiles of the generated MLEs. The distributions of the estimates from the precise data are plotted in Figure 2.7.

The distributions of $\hat{\lambda}$ and especially $\hat{\alpha}$ are skewed even before interval-censoring is introduced; this indicates a need for an alternative approximate distribution for the sampling distributions of $\hat{\alpha}$ and $\hat{\lambda}$. We also examine the distributions of $\hat{\lambda}$ and $\hat{\alpha}$ based on the generated interval-censored data with $w_i = 1 \; \forall \; i$, and then $w_i = 10 \; \forall \; i$, in Figure 2.8.

It is not surprising that the sampling distributions of the estimators from the interval-censored data with $w_i = 1 \; \forall \; i$ have a shape very similar to that of the precise

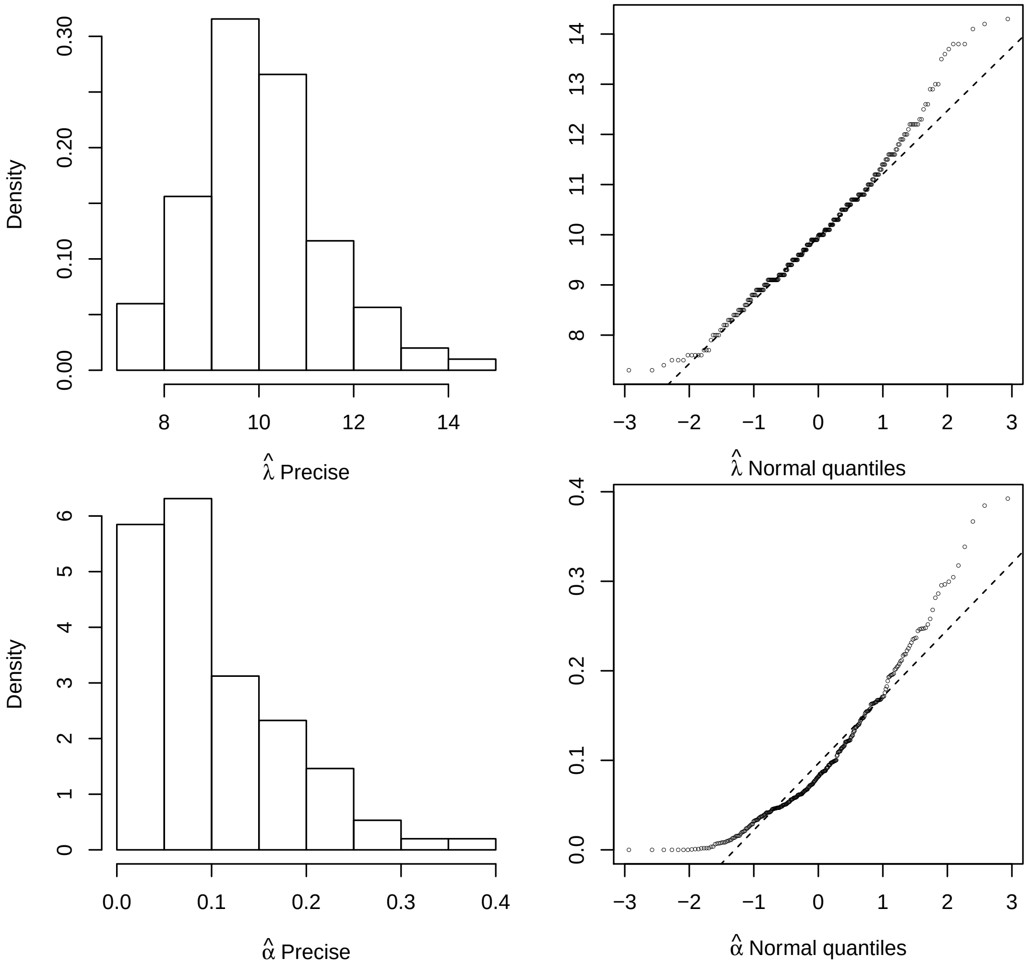

Figure 2.7: Histograms (left) and normal quantile plots (right) of the empirical sampling distributions of $\hat{\lambda}$ (top panels) and $\hat{\alpha}$ (bottom panels), precise data case.

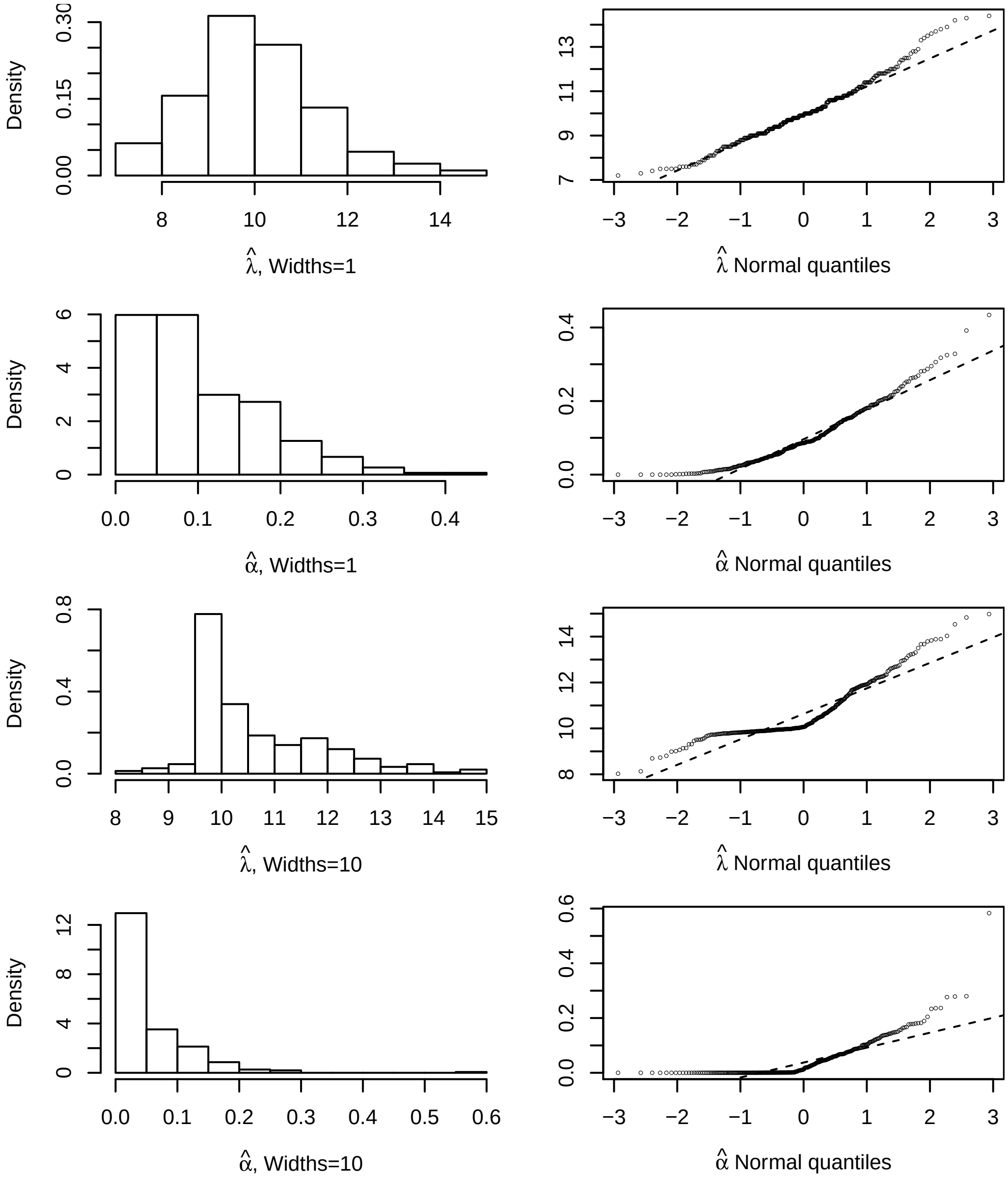

Figure 2.8: Histograms (left) and normal quantile plots (right) of the empirical sampling distributions of $\hat{\lambda}$ and $\hat{\alpha}$ based on 300 replications of samples of size $n = 10$ for the interval-censored data case, with widths = 1 (top four panels) and widths = 10 (bottom four panels).

data in Figure 2.7. There is a definite shift in the shape of the distribution of the estimates when we increase to $w_i = 10 \; \forall \; i$. The distribution of the $\hat{\lambda}$'s becomes more skewed, and the $\hat{\alpha}$'s tend more toward zero and also increase in skewness. Figure 2.8 further demonstrates that, for small n, the normal approximation to the sampling distributions of $\hat{\lambda}$ and $\hat{\alpha}$ can be quite poor, and thus the use of the Wald interval may not be appropriate.

The Wald formula is based on the large sample normal approximation for the sampling distribution of $\hat{\theta}$ given by (2.9), for fixed θ. The approximate density for $\hat{\theta}$, given θ, is

$$p_\theta(\hat{\theta}) \approx (2\pi)^{-1/2} |I(\hat{\theta})|^{1/2} \exp\left\{ -\frac{I(\hat{\theta})}{2} (\hat{\theta} - \theta)^2 \right\}, \qquad (2.10)$$

where $I(\hat{\theta})$ is the observed Fisher information. We calculate this approximation for the sampling distributions of $\hat{\lambda}$ and $\hat{\alpha}$ for the precise, interval-censored with $w_i = 1 \; \forall \; i$ and $w_i = 10 \; \forall \; i$. Note that we have set θ equal to the value used in generating the data. This could not be done in practice, of course, but (2.10) is only shifted in location as a function of θ, and not changed in shape. In Figure 2.9 we superimpose the approximations over the histograms from Figures 2.7 and 2.8.

The quadratic approximation appears to be a good fit to the sampling distribution in the case of $\hat{\lambda}$ for the precise and mildly censored data. The fits are less appropriate for the sampling distributions of the $\hat{\alpha}$'s, which are highly skewed, as well as the skewed sampling distribution of $\hat{\lambda}$ in the heavily censored case.

In the simulation experiments presented in Section 2.7, we compare Bayesian interval estimates to Wald estimates, but, as these examples demonstrate for a small-sample scenario, this approximation is not always appropriate. Ultimately we are concerned with building Bayesian models for interval-censored count data, but we include Wald interval estimates in our comparisons for completeness. We do not suggest to routinely use Wald interval estimates in the frequentist approach.

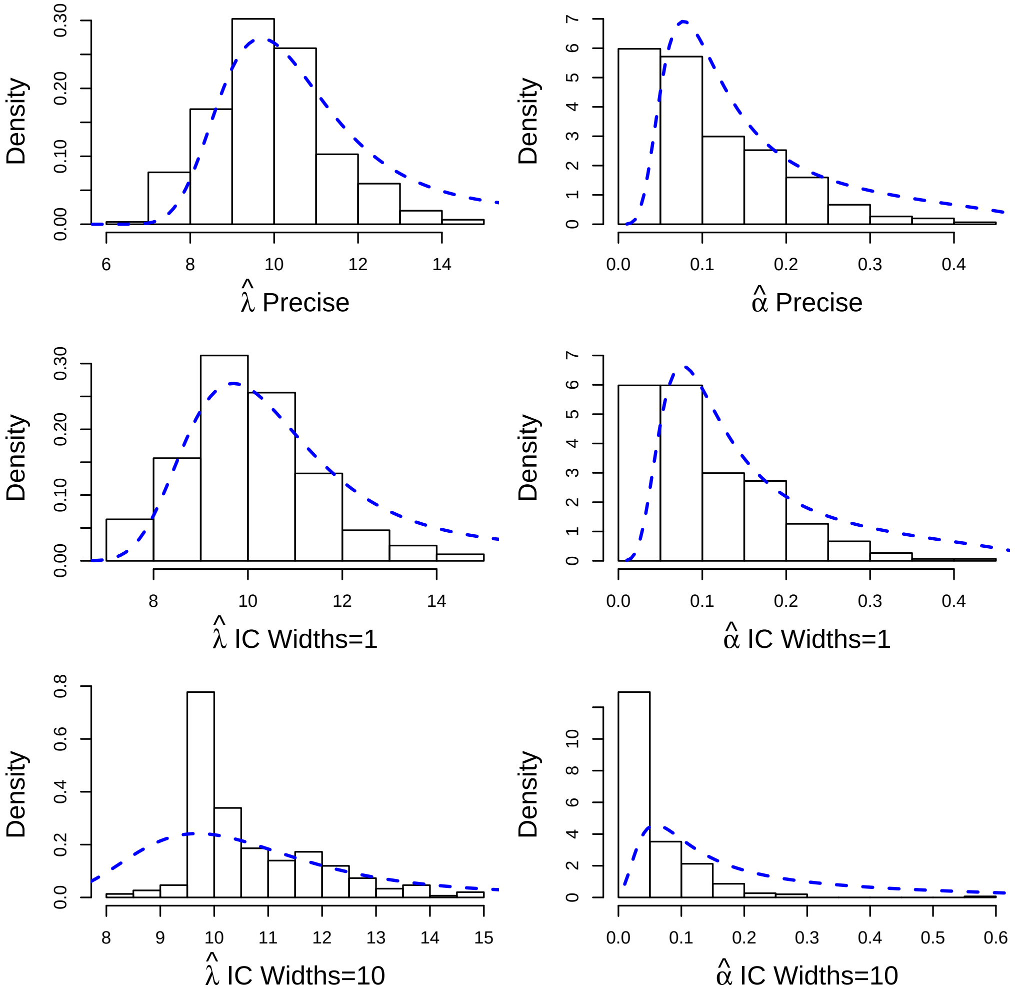

Figure 2.9: Histograms of maximum likelihood estimators based on 300 replications of samples of size $n = 10$ with normal theory approximations to the corresponding densities superimposed (blue dashed line). The left set of panels are for estimators $\hat{\lambda}$; the right for estimators $\hat{\alpha}$. The top row of panels is for precise data; the middle row for interval-censored data with $w = 1$; the bottom row for interval-censored data with $w = 10$.

2.5 Bayesian Development

As with any Bayesian model we need to specify two components: the likelihood function and the prior distribution. For our interval-censored problem there are advantages and disadvantages to various parameterizations. In this section we will examine some possible parameterizations.

Our choices in prior distributions need to reflect the prior beliefs on the unknown parameters λ and α. One approach would be to place priors directly on these two parameters of interest. A gamma distribution is suitable for each prior, as both parameters must be nonnegative. We let $\lambda \sim \text{gamma}(\nu_1, \nu_2)$ and, independently, $\alpha \sim \text{gamma}(\gamma_1, \gamma_2)$ so that the posterior distribution is

$$\pi(\lambda, \alpha | j \leq y \leq k) \propto \left[\sum_{y=j}^{k} \frac{\Gamma(y + 1/\alpha)}{y! \Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\lambda} \right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha\lambda} \right)^{y} \right] \\ \times \left[\nu_2^{\nu_1} \frac{1}{\Gamma(\nu_1)} \lambda^{\nu_1 - 1} e^{-\nu_2 \lambda} \right] \left[\gamma_2^{\gamma_1} \frac{1}{\Gamma(\gamma_1)} \alpha^{\gamma_1 - 1} e^{-\gamma_2 \alpha} \right].$$

This independent parameterization may be problematic, as we do not take into account the relationship between α and λ. Alternatively we may place independent priors on r and π, then the induced priors on α and λ will be dependent on each other because $\alpha = 1/r$ and $\lambda = r(1-\pi)/\pi$. We assign a prior distribution to r rather than α because it is the default parameter to use in the *WinBUGS* parameterization of the negative binomial distribution. If we place a gamma distribution on r then the prior induced on $\alpha = 1/r$ has an inverse gamma distribution where, if $r \sim \text{gamma}(a, b)$ then $\alpha = 1/r \sim \text{IG}(a, 1/b)$. Thus, because $\lambda = (1-\pi)/\alpha\pi$, a dependence between λ and α is forced in generating prior values. Because π is a probability, it is natural to assign a beta distribution to this parameter. Then, for $\pi \sim \text{beta}(\delta_1, \delta_2)$ and $r \sim \text{gamma}(\gamma_1, \gamma_2)$, this parameterization yields the posterior

$$\pi(\lambda,\alpha|j \leq y \leq k) \propto \left[\sum_{y=j}^{k} \frac{\Gamma(y+r)}{y!\Gamma(r)}\pi^r(1-\pi)^y\right] \times \left[\frac{1}{B(\delta_1,\delta_2)}\pi^{\delta_1-1}(1-\pi)^{\delta_2-1}\right]\left[\gamma_2^{\gamma_1}\frac{1}{\Gamma(\gamma_1)}r^{\gamma_1-1}e^{-\gamma_2 r}\right].$$

This will be our preferred parameterization in the Bayesian analysis.

2.5.1 Problems with Concentrated Gamma Priors

One commonly used “diffuse gamma prior” for a precision parameter is the gamma(0.001,0.001) distribution. If, for example, when sampling from a normal distribution mean μ and precision τ (both unknown) placing a normal prior on $\mu|\tau$ and a gamma$(0.001, 0.001)$ prior on τ will yield a very diffuse joint prior, albeit proper and conjugate. This is because the dispersion of a normal distribution is determined by its variance, $1/\tau$, and the gamma$(0.001, 0.001)$ is itself highly concentrated, placing most of its probability mass near zero. If we do not have background knowledge about our parameters, we may be tempted to set priors $\pi \sim$ beta$(1, 1)$ and $r \sim$ gamma$(0.001, 0.001)$, with the goal of having the data dominate the posterior. This prior structure is problematic, however. Placing a gamma(0.001,0.001) prior on a parameter is indicating heavy prior belief on a value near zero for that parameter (a “spike” near zero). This spike near zero for r in turn affects the priors on α and λ. We illustrate the danger in using such a prior in Figure 2.10. We use our $n = 10$, interval-censored data set with $w_i = 5 \ \forall \ i$ from Figure 2.3 and plot the joint prior for (λ, α) based on $(\pi, r) \sim$ beta(1,1)×gamma(0.001,0.001), interval-censored likelihood, and resulting posterior.

Note that the “diffuse” gamma priors have actually influenced the posterior, placing heavy probability at zero for r causes higher prior probabilities on higher values for α, and this in turn pulls the posterior for α in the positive direction, away from the true value 0.1. This effect is more visually evident in the profile likelihood and marginal posterior for α, plotted together on the right side of Figure 2.11.

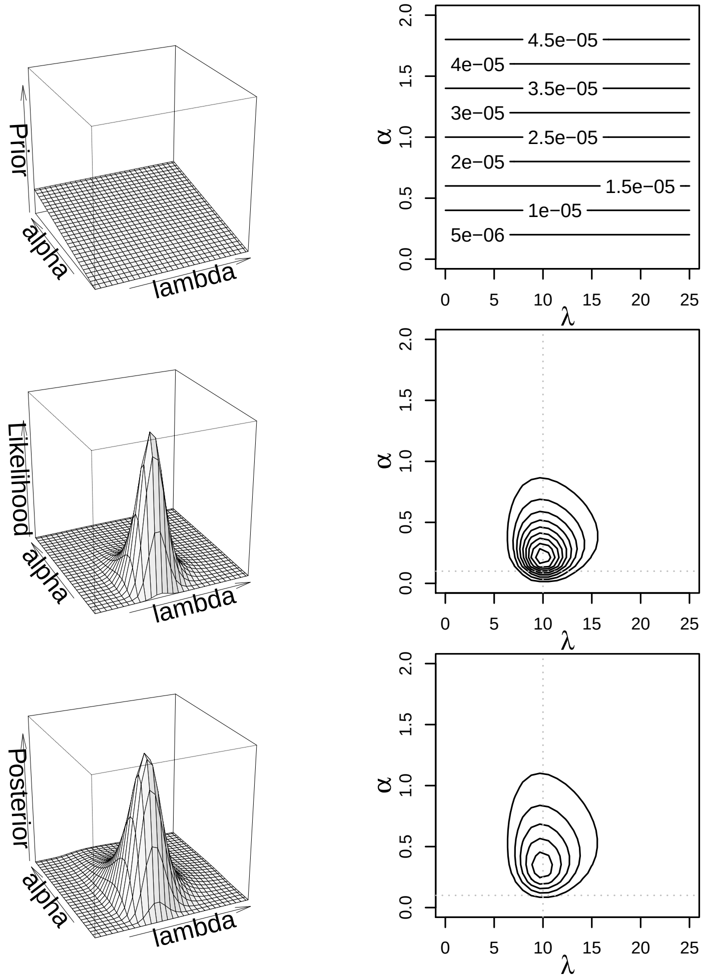

Figure 2.10: Joint prior (top left panel) based on $r \sim$ gamma(0.001,0.001) and $\pi \sim$ beta(1,1), interval-censored $w = 5$ likelihood (center left panel), and resulting posterior (bottom left panel) and their corresponding contour plots (right panels).

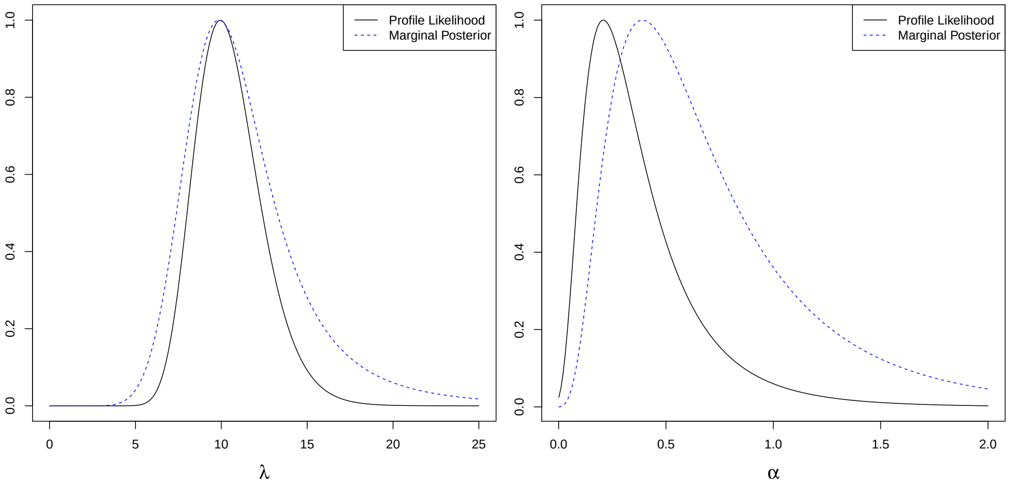

Figure 2.11: Profile likelihoods (solid black curves) and marginal posteriors (dashed blue curves) for λ (left panel) and α (right panel).

Using a less concentrated gamma prior would avoid these problems. In Figure 2.12 we generate $n = 10$ precise observations from a NB(10,0.1) distribution. The center panel is the contour plot of the likelihood. We also generate interval-censored data, with $w_i = 5 \ \forall \ i$, based on the precise values, and the contours of the interval-censored likelihood are superimposed in red. Our first prior structure uses $\pi \sim \text{beta}(1,1)$ and the suggested $r \sim \text{gamma}(2,2)$. The contours of the prior and resulting posterior on λ and α are plotted on the top row in Figure 2.12. We also show the interval-censored posterior contours in red.

Note that the gamma(2,2) prior structure on r inflates α. Furthermore, in the λ direction, this prior increases posterior dispersion compared to the likelihood. For an alternative prior structure in a further attempt to keep the posterior shape similar to the likelihood (keeping the prior "non-informative" relative to the likelihood), we again change the gamma prior on r. It should be noted that our comparison of the likelihood to the posterior in these examples is for purposes of illustration only: we are attempting to demonstrate the effect of informativeness in prior structures. In practice, the prior would not be selected by comparison to the likelihood, though prior-to-posterior robustness is an important issue in Bayesian modeling.

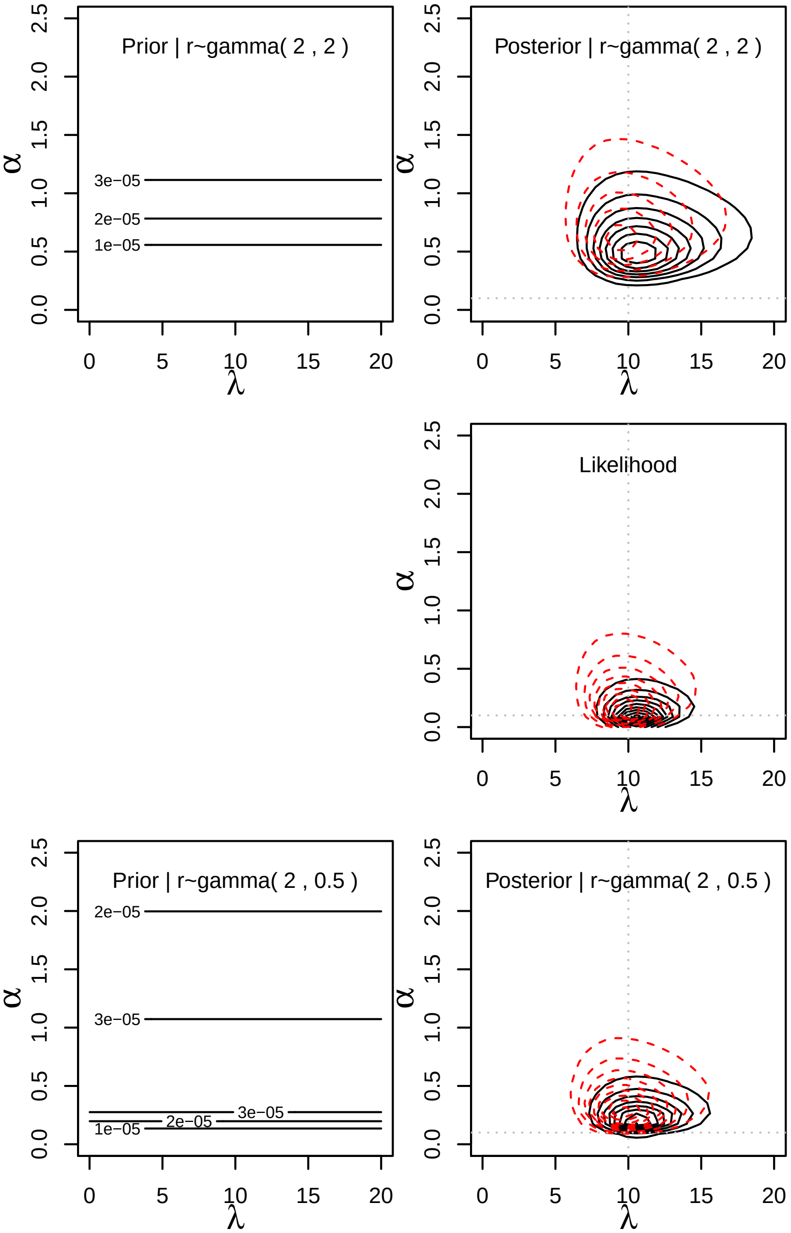

Figure 2.12: Joint prior and resulting posterior on (λ, α) for r ~gamma$(2, 2)$ (top panels), based on likelihood from $n = 10$ observations from a NB(10,0.1) (center panel), and prior with resulting posterior for r ~gamma$(2, 0.5)$ (bottom panels). The precise (solid black) and interval-censored with $w = 5$ (dashed red) contours are given for the likelihood and posteriors.

We base our next choice on the behavior of the induced prior on α, as that is our dispersion parameter of interest. When we place a gamma(2,2) on r, the corresponding distribution on α is inverse gamma IG(2,1/2). These densities are plotted on the left in Figure 2.13. The effect of the relatively diffuse prior on r is to create a spike near zero in the density for α. Taking $\alpha \sim \text{gamma}(2, 1/2)$ yields similar results for the posterior, as can be seen in the right-hand plot in Figure 2.13.

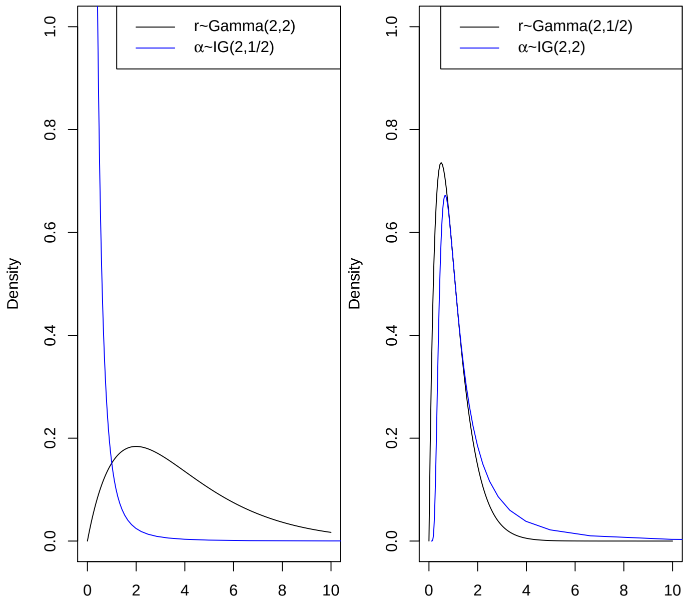

Figure 2.13: Prior choices on r (plotted in black) of gamma(2,2) distribution (left panel) and gamma(2,1/2) distribution (right panel) with corresponding induced priors on α (plotted in blue).

The bottom two contour plots in Figure 2.12 make use of this alternative gamma$(2, 1/2)$ prior on r. Note that the posterior contours are more similar to the original likelihood than those obtained using a gamma(2,2) on r.

Next we consider data generated with $\lambda = 10$ as before, but increasing α to 1. We again generate $n = 10$ observations, this time from a NB(10,1) distribution, and we again employ first a gamma(2,2) and then a gamma(2,1/2) prior on r. The results are displayed in Figure 2.14.

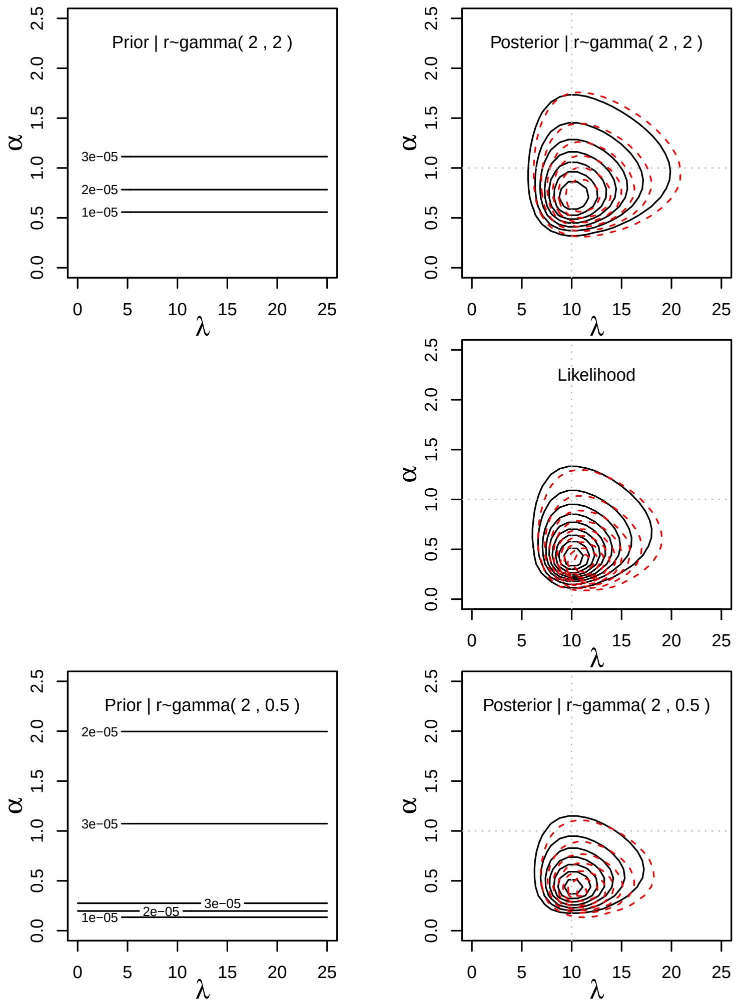

Figure 2.14: Joint prior and resulting posterior on (λ, α) for r ~gamma$(2, 2)$ (top panels), based on likelihood from $n = 10$ observations from a NB(10,1) (center panel), and prior with resulting posterior for r ~gamma$(2, 0.5)$ (bottom panels). The precise (solid black) and interval-censored with $w = 5$ (dashed red) contours are given for the likelihood and posteriors.

This time, using the gamma(2,2) prior on r actually leads to posterior results closer to the original likelihood than using the gamma(2,1/2). This situation is reversed from what we saw in Figure 2.12, where the gamma(2,1/2) prior led to more a more similar posterior. For a final example we again increase the true value of α to 2. The results are plotted in Figure 2.15.

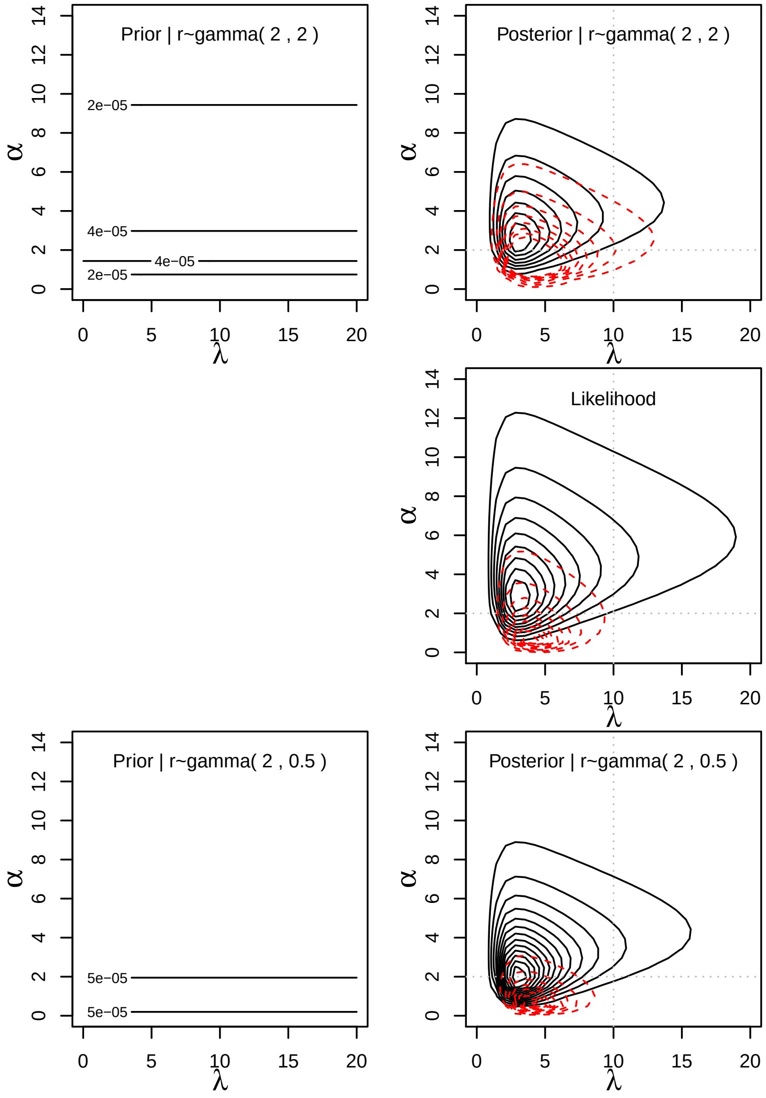

Figure 2.15: Joint prior and resulting posterior on (λ, α) for r ~gamma$(2, 2)$ (top panels), based on likelihood from $n = 10$ observations from a NB(10,2) (center panel), and prior with resulting posterior for r ~gamma$(2, 0.5)$ (bottom panels). The precise (solid black) and interval-censored with $w = 5$ (dashed red) contours are given for the likelihood and posteriors.

For this likelihood situation we see that neither originating prior structure leads to a posterior that comes very close in shape to the original likelihood, although $r \sim \text{gamma}(2, 0.5)$ seems to do a slightly better job. Figures 2.12, 2.14, and 2.15 illustrate how difficult it will be, at least in the small-sample case, to find a general prior structure on r that is sufficiently "non-informative" regardless of the originating priors on α and λ.

It is possible to alleviate issues with induced priors by using larger sample sizes; Figure 2.16 demonstrates the effect when we increase n to 50. The likelihood contour represents 50 observations generated from a NB(10,1) and corresponding interval-censored values of width 5 plotted in red. In this case, regardless of the choice of gamma prior on r, the posterior contours resemble those of the original likelihood. This suggests that the sample size was sufficiently large to overpower the effect of the four prior choices: $r \sim \text{gamma}(2, 2)$, $r \sim \text{gamma}(2, 0.5)$, $r \sim \text{gamma}(0.001, 0.001)$, and $r \sim \text{gamma}(1, 1)$. We find this to be true for several generated datasets with $\alpha = 1$, as well as with $\alpha = 2$. When we decrease α to 0.1, a larger sample size is required to overpower prior influence—about 100 observations.

2.5.2 Uniform Prior with Large Upper Bound

We consider a uniform prior on r as an alternative to the suggested gamma prior. Spiegelhalter *et al.* (2004, pp. 168-177) suggest a $\text{uniform}(0, B)$ for a dispersion parameter, where the upper bound B is chosen so as to stabilize posterior features of interest, such as credible interval width or a specified posterior probability. To apply this prior to various data sets we might simply choose B to be large, say, $r \sim \text{uniform}(0, 1000)$. To illustrate this prior, we generate new precise data sets of size 10 from a $\text{NB}(10, 0.1)$, $\text{NB}(10, 1)$ and $\text{NB}(10, 2)$, with corresponding interval-censored data with $w_i = 5 \; \forall \; i$ as in Section 2.5.1. In Figures 2.17, 2.18 and 2.19, we plot the joint prior on α and λ based on $\pi \sim \text{beta}(1, 1)$ and $r \sim \text{uniform}(0, 1000)$,

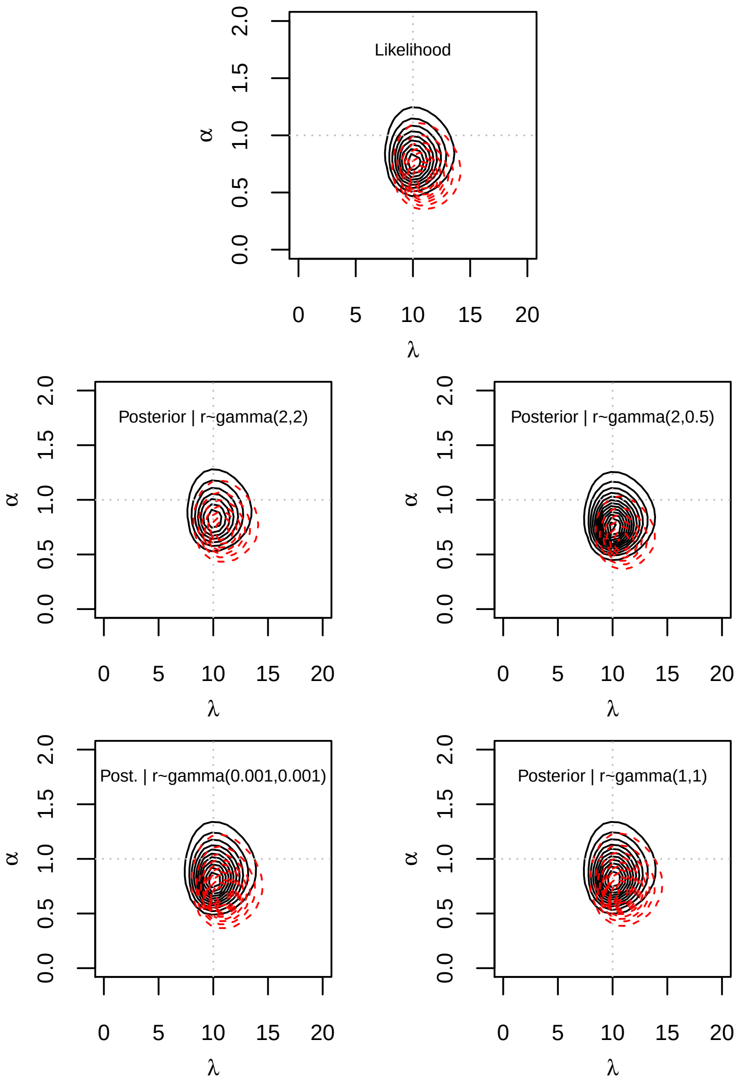

Figure 2.16: Likelihood contours for $n = 50$ observations from a NB(10,1) distribution (top panel) and corresponding posterior contours based on $\pi \sim \text{beta}(1, 1)$ and four choices of prior structure for r: gamma(2,2) (center left panel), gamma(2,1/2) (center right panel), gamma(0.001,0.001) (bottom left panel), and gamma(1,1) (bottom right panel). Contours in each plot are given for precise case (solid, black) and interval-censored case with $w = 5$ (dashed, red).

together with the likelihoods and posteriors for each of the three data scenarios. As before, the interval-censored results are in red.

The posteriors seem to shrink from the likelihoods in each case (data based on $\alpha = 0.1$, 1, or 2) but the locations of the posteriors relative to the likelihoods remain the same—this is in contrast to, for example, the effect of the gamma(2,2) prior on the posteriors in Figures 2.12 and 2.14. In Figure 2.15 the gamma(2,2) and the gamma(2,0.5) both result in a posterior that maintains location relative to the likelihood, but the shrinkage effect is more drastic than that of the uniform in Figures 2.17, 2.18 and 2.19. Based on these observations, we will use $r \sim \text{uniform}(0, 1000)$ in our simulation experiments in Section 2.7. It should be noted, however, that we are not proposing the uniform(0,1000) as an automatic prior diffuse prior for the dispersion parameter. For example, use of this prior in our more complex regression models has led to convergence issues.

2.6 Bayesian Analysis Examples

Once we have determined our prior parameterization, we use Markov chain Monte Carlo (MCMC) methods to obtain posterior distributions given the lack of closed-form solutions. We revisit the examples introduced in Section 2.3 in order to introduce the Bayesian approach to the analysis. In Section 2.7, we present results based on full simulation studies.

For all of the examples in this section, we use the priors $\pi \sim \text{beta}(1, 1)$ and $r \sim \text{gamma}(2, 2)$. The data are generated from a negative binomial distribution with specified mean and dispersion parameter. In each example, we run the *WinBUGS* analysis first for the precise data using a burn-in of 5000, thinning rate of 30 and 5000 updates (after thinning from a total chain length of 150000). This yields posterior estimates for both α and λ.

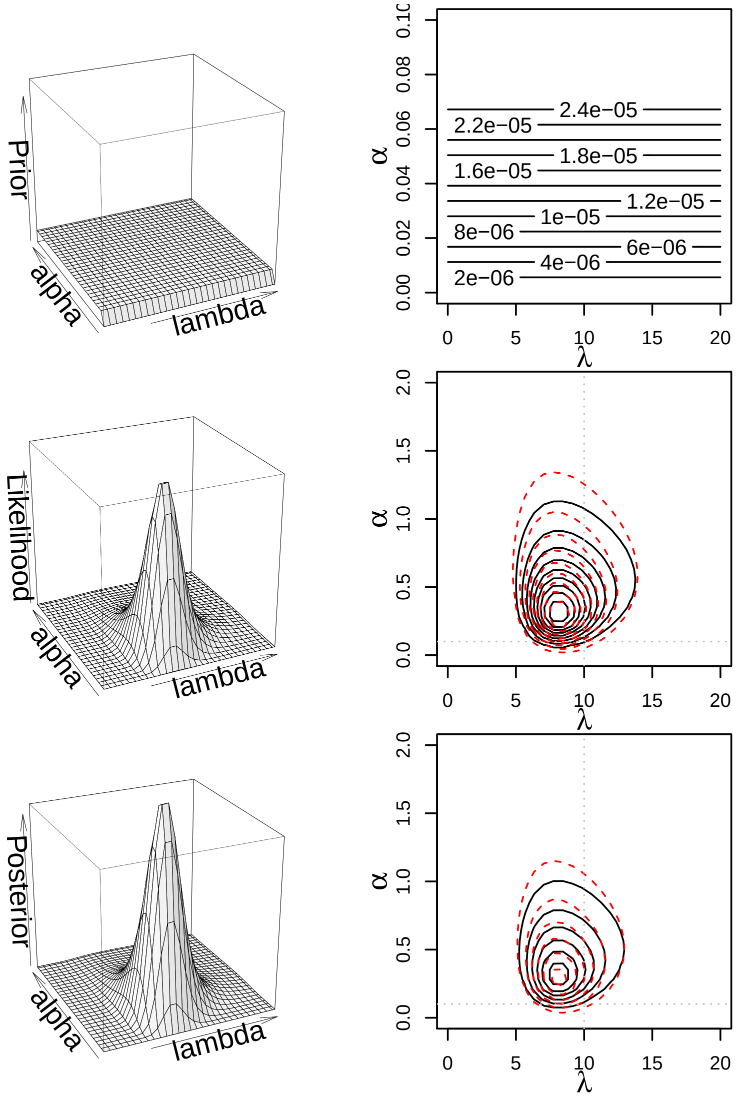

Figure 2.17: Joint prior (top left panel) based on r ∼uniform(0,1000) and π ∼beta(1,1), precise likelihood from a NB(10,0.1) (center left panel), and resulting posterior (bottom left panel) and their corresponding contour plots (right panels). In addition to the precise case contours (solid black), the $w = 5$ interval-censored contours (dashed red) are given.

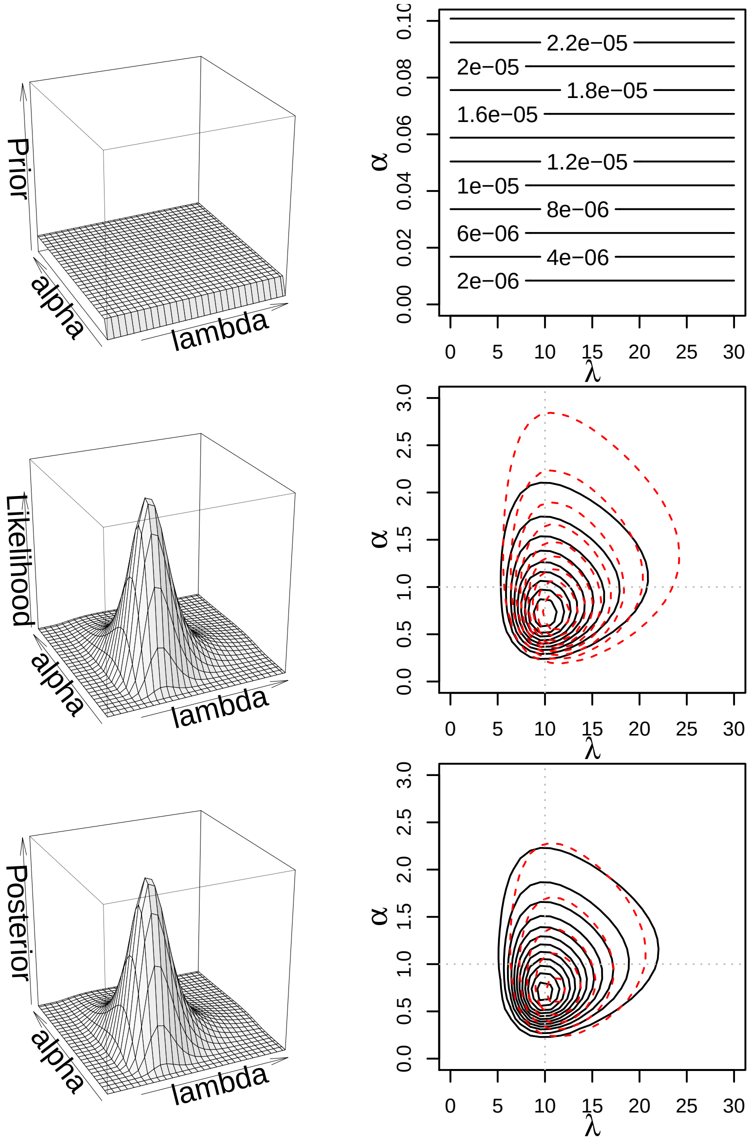

Figure 2.18: Joint prior (top left panel) based on r ∼uniform(0,1000) and π ∼beta(1,1), precise likelihood from a NB(10,1) (center left panel), and resulting posterior (bottom left panel) and their corresponding contour plots (right panels). In addition to the precise case contours (solid black), the $w = 5$ interval-censored contours (dashed red) are given.

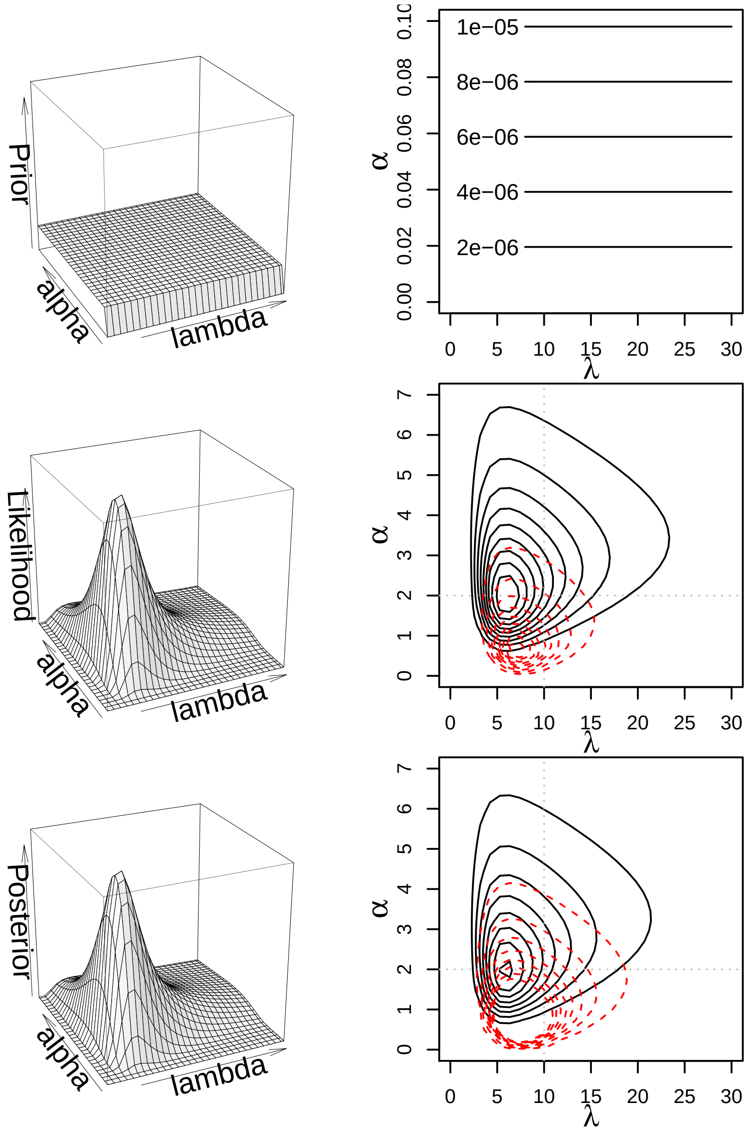

Figure 2.19: Joint prior (top left panel) based on r ∼uniform(0,1000) and π ∼beta(1,1), precise likelihood from a NB(10,2) (center left panel), and resulting posterior (bottom left panel) and their corresponding contour plots (right panels). In addition to the precise case contours (solid black), the $w = 5$ interval-censored contours (dashed red) are given.

To implement the interval-censored likelihood (2.4) in *WinBUGS*, we must use the "zeros trick". See Spiegelhalter *et al.* (2003) for information on this method. We use chains of the same length and composition as in the precise case, and compare posterior features from each.

For our first example we use the data that formed the likelihoods in Figure 2.2. In that example we generated $n = 10$ exact counts from a $NB(10, 0.1)$ distribution to form $\mathbf{y}$, and each interval width in the interval-censored data set was only 1 unit long.

Convergence diagnostics for this *WinBUGS* run are given in Appendix A. The thinning rate of 30 produces sufficient dampening in the autocorrelation plots, and there is satisfactory mixing of the two chains. Subsequent examples in this section produced similar convergence diagnostics. The posteriors for λ and α under each method are plotted in Figure 2.20. Results are as expected given the likelihoods in Figure 2.2 and this relatively diffuse prior structure. The posteriors for α and λ are virtually identical for the two methods. The plot on the right of Figure 2.20 represents draws from the two joint posteriors (blue for precise and red for interval-censored) as well as a sample from the prior plotted in black. The posterior summaries for the two methods are given in Table 2.2.

Table 2.2. Bayesian Estimates for Precise and Interval-Censored Data

Parameter	True	Prec	Prec	Prec	IC	IC	IC
	Value	Mean	S.D.	95% C.S.	Mean	S.D.	95% C.S.
λ	10	9.36	2.38	(5.64,14.86)	9.28	2.38	(5.63,14.83)
α	0.1	0.51	0.21	(0.24,1.04)	0.51	0.21	(0.24,1.04)

Next we consider the effect of increasing the widths of the intervals in the interval-censored data set, using the same generated data from Figure 2.3. As in Section 2.3, we consider $w_i = 5$ and $w_i = 10 \,\forall\, i$, in turn. The precise data set remains

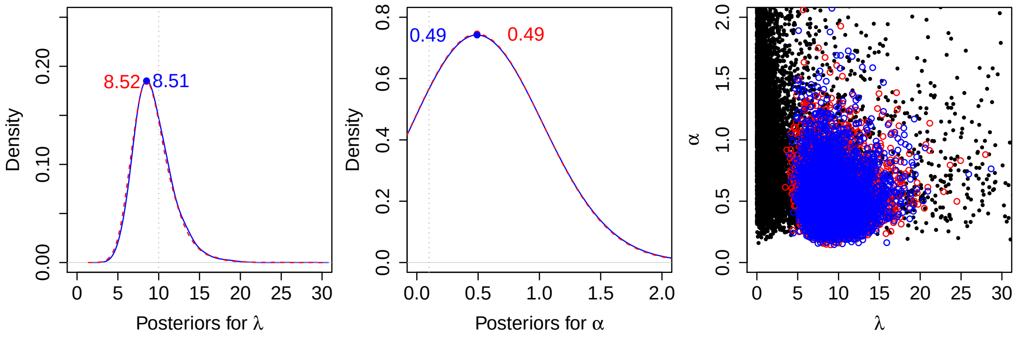

Figure 2.20: Bayesian analysis of data represented in Figure 2.2. Precise (solid, blue) and $w = 1$ interval-censored (dashed, red) marginal posteriors are plotted for λ (left panel) and α (center panel). The right panel illustrates draws from the joint posterior for λ and α for the precise (blue) and interval-censored (red) cases, as well as a random sample from the joint prior (black).

unchanged in each case. As expected, the posterior standard deviations increase with increasing interval widths. The posterior for λ exhibits more variability when $w_i = 5$, and the posteriors for both λ and α present more variability than the precise case in the bottom row, when $w_i = 10$. This is in agreement with what we saw for the pure likelihoods in Figure 2.3. The posterior summaries are given in Table 2.3.

Table 2.3. Bayesian Estimates: Increasing Interval-Censored Widths, $\lambda = 10$, $\alpha = 0.1$

Param	w	Mean	S.D.	95% C.S.	w	Mean	S.D.	95% C.S.
λ	Prec	9.36	2.38	(5.64,14.86)	1	9.28	2.38	(5.63,14.83)
α	Prec	0.51	0.21	(0.24,1.04)	1	0.51	0.21	(0.24,1.04)
λ	5	9.86	2.64	(5.81,16.09)	10	8.76	2.86	(4.32,15.57)
α	5	0.55	0.23	(0.25,1.11)	10	0.77	0.43	(0.29,1.90)

Note in Table 2.3 that all of the posterior means for α, from the precise case to the interval-censored with $w_i = 10\ \forall\ i$ case, over-estimate the true value of α from which the data were generated. This is expected given the effect seen in Figure 2.12 of using $r \sim \text{gamma}(2, 2)$ on data generated from a NB(10,0.1), similar to the data used here. As an alternative we run the analysis again using $r \sim \text{unif}(0, 1000)$ as discussed in Section 2.5.2. The new posterior results are given in Table 2.4. The

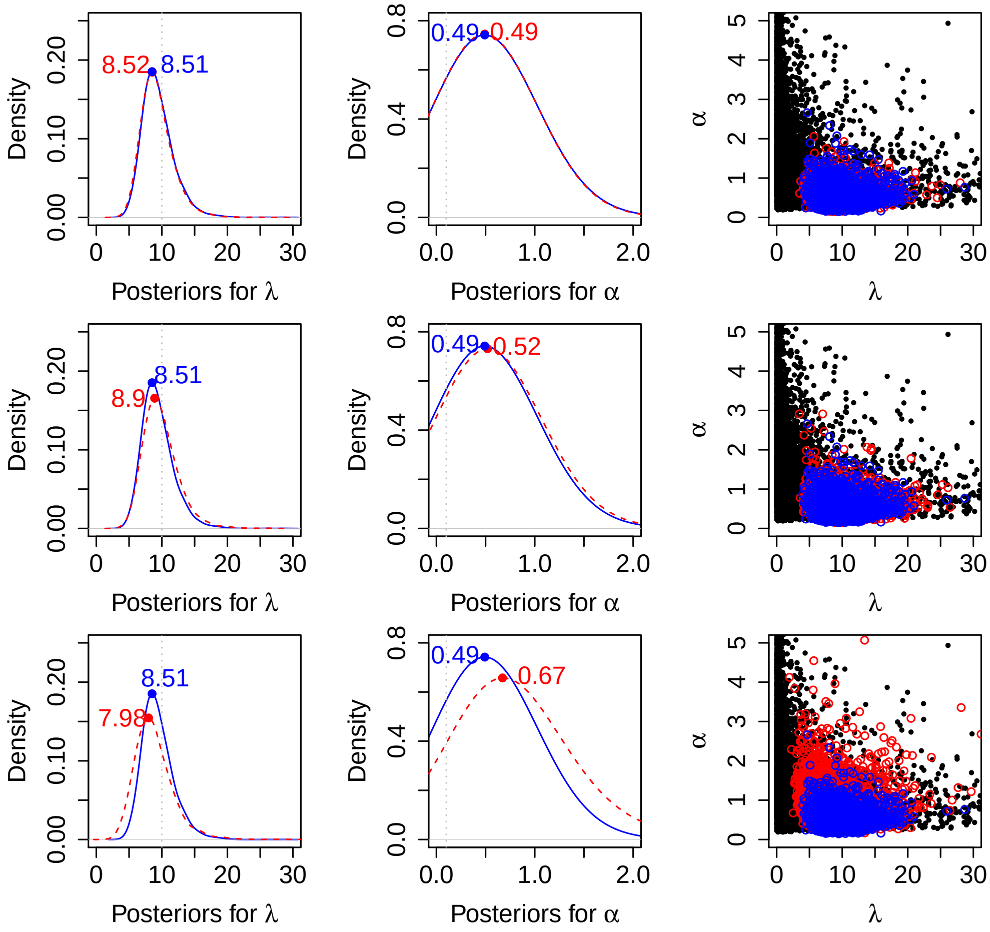

Figure 2.21: Bayesian analysis of data represented in Figure 2.3. Precise (solid, blue) marginal posteriors are plotted for λ (left column) and α (center column). The interval-censored posteriors (dashed, red) represent interval-censored data from $w = 1$ (top row) to $w = 5$ (center row) to $w = 10$ (bottom row). The right column illustrates draws from the joint posterior for λ and α for the precise (blue) and interval-censored (red) cases, as well as a random sample from the joint prior (black).

uniform prior on r dramatically changes the posterior means and interval estimates; all posterior means are now close to the true $\alpha = 0.1$ and all posterior credible sets are shifted downward. The posterior standard deviations for λ are also reduced.

Table 2.4. Estimates from $r \sim \text{unif}(0, 1000)$: Increasing IC Widths, $\lambda = 10$, $\alpha = 0.1$

Param	w	Mean	S.D.	95% C.S.	w	Mean	S.D.	95% C.S.
λ	Prec	9.41	1.42	(6.89,12.47)	1	9.32	1.38	(6.82,12.28)
α	Prec	0.12	0.10	(0.01,0.38)	1	0.11	0.10	(0.00,0.38)
λ	5	10.15	1.65	(7.21,13.59)	10	10.28	1.66	(7.05,13.51)
α	5	0.13	0.12	(0.01,0.43)	10	0.08	0.16	(0.00,0.44)

We now consider a Bayesian analysis of the data plotted in the contours in Figure 2.4. In that figure, increasing the percentage of data censored inflated estimates of λ and attenuated estimates of α, depending on where the intervals were positioned relative to the true values. As expected, the results are similar: Posterior means for λ are inflated, while those of α are slightly attenuated. The posterior summaries are given in Table 2.5.

Next we illustrate the effect of sample size, using the data that produced the likelihoods in Figure 2.5. The shapes of the posteriors are very similar at each sample size, with slight differences noted only for $n = 5$. In Table 2.6 we see that in the $n = 5$ case the posterior standard deviation is slightly higher for the interval-censored data than for the precise counts; for $n = 20$ and $n = 50$ the posterior variability is the same across the two methods.

We continue with the Bayesian analysis of the data in Figure 2.6, in which we shifted the intervals in their positioning relative to the precise data. These results reflect what we observed with the likelihoods in Figure 2.6. When the precise value tends toward the lower endpoint of each interval (top row), there is more bias in the posterior mean for λ and less bias in that for α. When the precise value is centered within each interval, the posteriors based on the precise data and on this type of

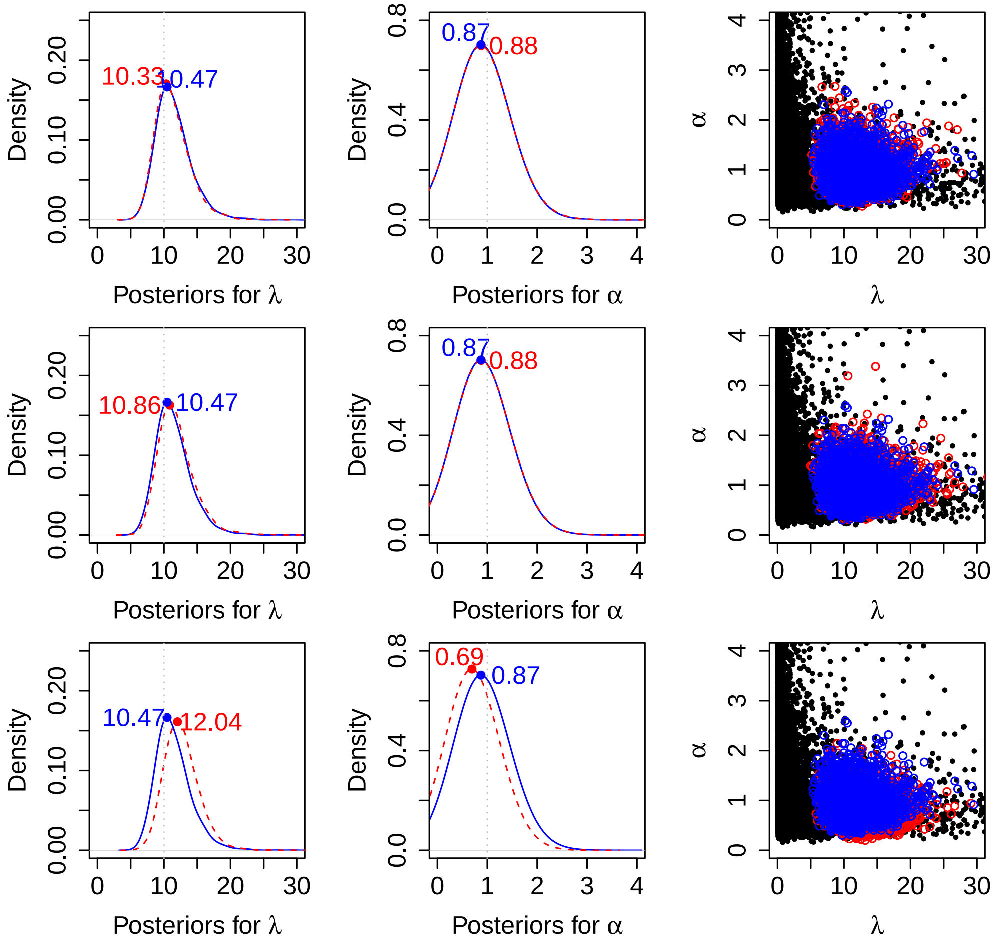

Figure 2.22: Bayesian analysis of data represented in Figure 2.4. Precise (solid, blue) marginal posteriors are plotted for λ (left column) and α (center column). The interval-censored posteriors (dashed, red) represent 5% interval-censored data (top row) to 50% interval-censored data (center row) to 100% interval-censored data (bottom row). The right column illustrates draws from the joint posterior for λ and α for the precise (blue) and interval-censored (red) cases, as well as a random sample from the joint prior (black).

Table 2.5. Bayesian Estimates: Increasing Percent of Censored Data, $\lambda = 10$, $\alpha = 1$

Param	% IC	Mean	S.D.	95% C.S.	% IC	Mean	S.D.	95% C.S.
λ	Prec	11.46	2.61	(7.35,17.50)	5%	11.31	2.56	(7.30,17.34)
α	Prec	0.90	0.28	(0.49,1.57)	5%	0.90	0.29	(0.48,1.59)
λ	50%	11.89	2.73	(7.68,18.23)	100%	12.66	2.56	(8.46,18.36)
α	50%	0.91	0.29	(0.49,1.60)	100%	0.71	0.24	(0.37,1.29)

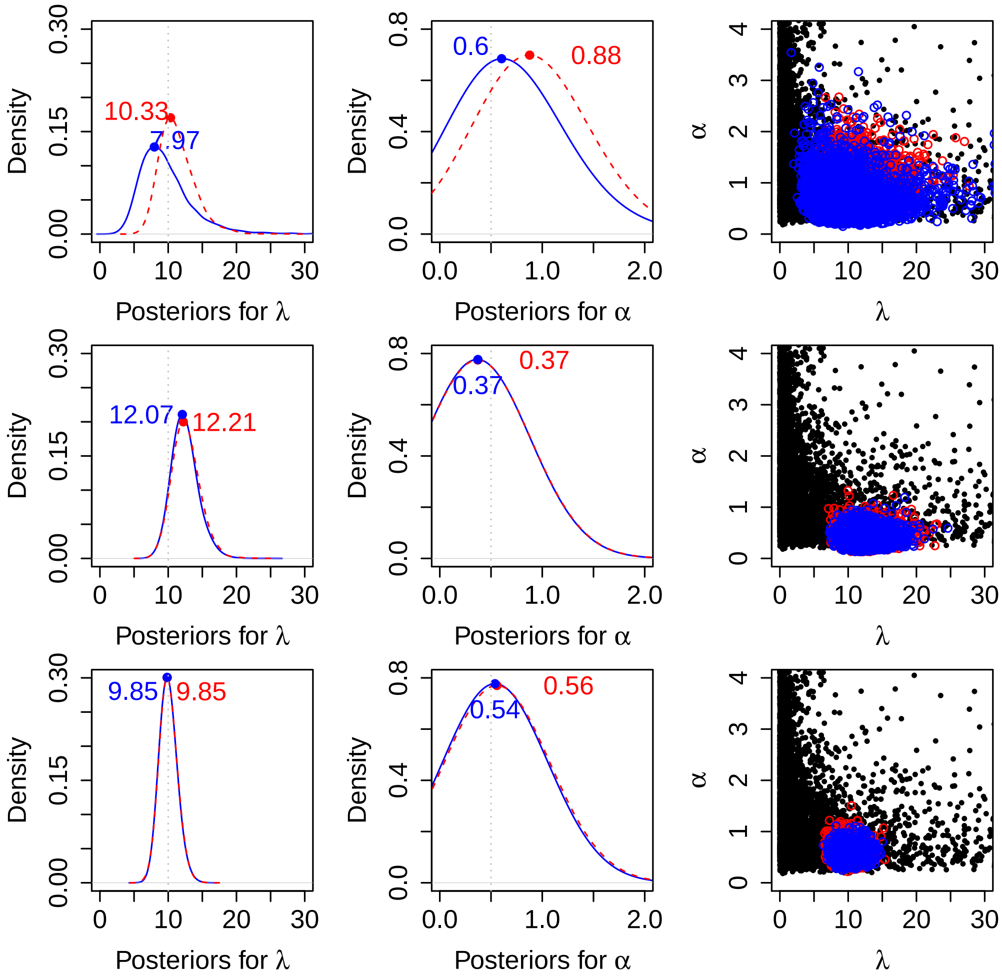

Figure 2.23: Bayesian analysis of data represented in Figure 2.5. Precise (solid, blue) and interval-censored with $w = 5$ (dashed, red) marginal posteriors are plotted for λ (left column) and α (center column). The sample size n increases from 5 (top row) to 20 (center row) to 50 (bottom row). The right column illustrates draws from the joint posterior for λ and α for the precise (blue) and interval-censored (red) cases, as well as a random sample from the joint prior (black).

Table 2.6. Bayesian Estimates: Increasing n, $\lambda = 10$ and $\alpha = 0.5$

Param	n	Prec Mean	Prec S.D.	Prec 95% C.S.	IC Mean	IC S.D.	IC 95% C.S.
λ	5	9.78	4.18	(4.43,19.91)	10.40	4.61	(4.47,21.70)
α	5	0.67	0.36	(0.26,1.56)	0.71	0.39	(0.27,1.69)
λ	20	12.47	1.89	(9.16,16.64)	12.60	1.96	(9.18,16.89)
α	20	0.37	0.12	(0.20,0.66)	0.37	0.12	(0.20,0.67)
λ	50	9.99	1.14	(8.00,12.52)	10.02	1.18	(7.94,12.52)
α	50	0.54	0.12	(0.36,0.80)	0.56	0.14	(0.34,0.88)

interval-censored data are very similar. When the precise value tends toward the upper endpoint of each interval (bottom row), the bias trend is reversed from the top case: the posterior mean for λ is less biased, but the posterior mean for α is more biased. The posterior summaries in Table 2.7 confirm these results.

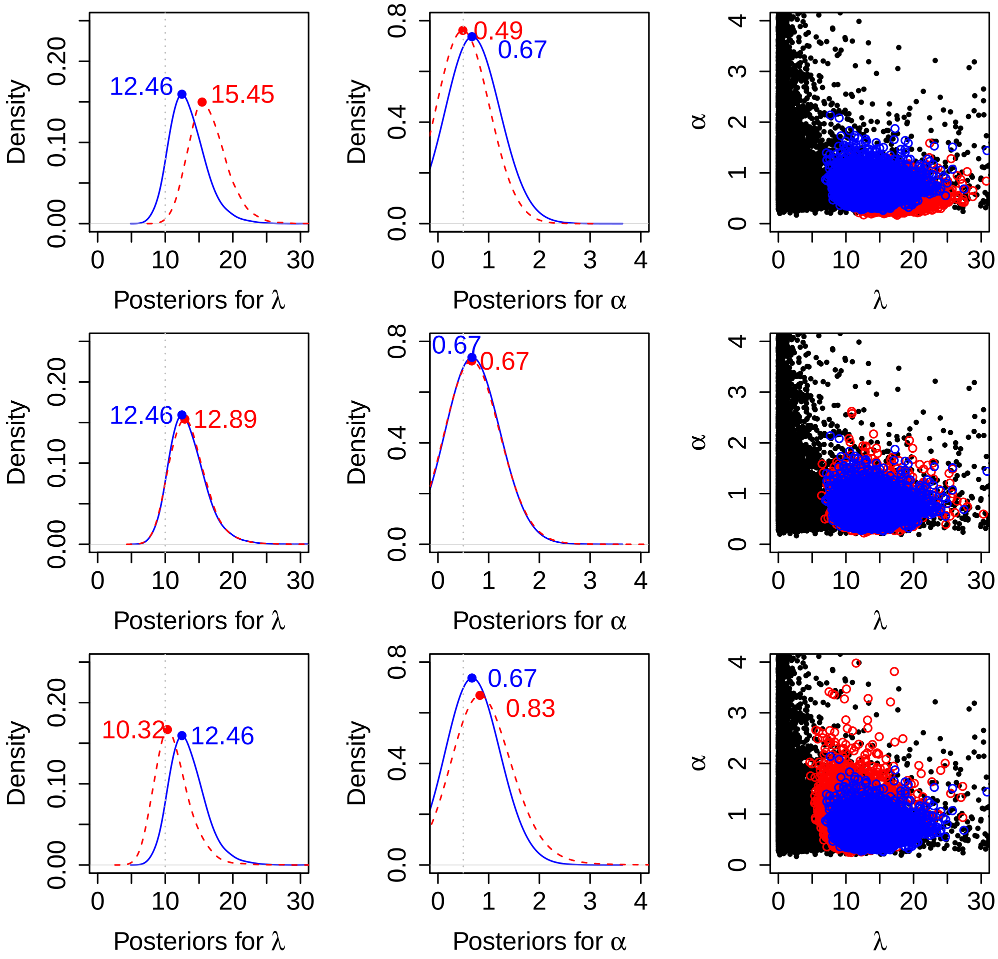

Figure 2.24: Bayesian analysis of data represented in Figure 2.6. Precise (solid, blue) and interval-censored with $w = 5$ (dashed, red) marginal posteriors are plotted for λ (left column) and α (center column). The right column illustrates draws from the joint posterior for λ and α for the precise (blue) and interval-censored (red) cases, as well as a random sample from the joint prior (black). The top row represents $j = y$, the middle row represents y centered in each interval, and the bottom row represents $k = y$.

Table 2.7. Bayesian Estimates: Changing Tendency of y, $\lambda = 10$ and $\alpha = 0.5$

Par	Tend	Mean	SD	95% CS	Tend	Mean	SD	95% CS
λ	Prec	13.39	2.68	(9.07,11.49)	j	16.36	2.78	(11.70,22.67)
α	Prec	0.69	0.21	(0.38,1.20)	j	0.50	0.16	(0.27,0.89)
λ	Center	13.22	2.79	(8.63,19.46)	k	11.11	2.63	(7.01,17.01)
α	Center	0.77	0.27	(0.38,1.44)	k	0.89	0.37	(0.40,1.80)

2.7 Simulation Studies

In this section, we present the results of simulation experiments designed to explore the themes we have seen in the examples. The simulations described below were implemented using *R2WinBUGS* on a high-performance computing system. Details of the computing system used for the simulations appear in Appendix B, and code is available upon request.

In Sections 2.3 and 2.6, we considered examples exhibiting several interesting features of our interval-censored problem. These included the widths of the intervals, the percent of the data that are interval-censored, the sample size of the data, and the positioning of the precise count within each interval. In the simulations described below, we study several of these scenarios. In these simulations, we focus on 4 distinct methods of handling interval-censored count data. These are as follows:

(1) Apply the interval-censored likelihood (2.4) to the data

(2) Fit the precise negative binomial model (2.2) using the "mean method": use the midpoint of each $[j_i, k_i]$, rounded to the nearest whole number, as the precise count

(3) Fit the precise negative binomial model (2.2) taking $y_i \equiv j_i \; \forall \; i$

(4) Fit the precise negative binomial model (2.2) taking $y_i \equiv k_i \; \forall \; i$

For each design point with specified λ, α, n, and interval width w, we generate $\mathbf{y}$ from a NB(λ, α) distribution. We fix w within each data set for simplicity. Next we generate interval-censored data based on our precise samples: for each y_i we randomly select its position p_i within the interval of width w. For example, if the interval width is $w = 2$ then p_i could equal 1, 2, or 3. Then we set $j_i = y_i - p_i$ and $k_i = y_i + w_i - p_i$. This is repeated for each data point in $i = 1, \ldots, n$.

Experimental design point selection involves several considerations. First, we select a sample size of $n = 50$ for each design point. We then need to select the true expected count, λ, from which we generate the data. We choose $\lambda = 3$ to generate counts close to zero, $\lambda = 10$ to generate counts in the intermediate range, and $\lambda = 30$ to generate larger counts. To select the experimental true value of α, we must consider the corresponding variability in the data that the choice of α will induce. Recall from Section 2.1 that the variance of the counts equals $\lambda + \alpha\lambda^2$. If $\lambda = 30$ and $\alpha = 1$, this corresponds to a variance of 930. For reasonable interval censoring, it may be difficult to detect differences in the precise versus the interval-censored data when the data as a whole are so dispersed. For $\lambda = 30$ we use widths of 1, 5, and 10 representing mild, intermediate, and heavy censoring, and for $\lambda = 10$ we select widths of 1, 3, and 5. Then we select α so as to induce a standard deviation in the data that will fall below the widest censoring width. For $\lambda = 3$ we consider only $w_i = 2 \; \forall \; i$, and we select $\alpha = 3$. This induces a standard deviation of 5.5, which does not fall below w but serves to demonstrate a case where α is larger but overall variance is still reasonable. These design points are summarized in Table 2.8.

Table 2.8. Simulation values of λ and α

λ	w	α	Var	Std Dev
3	2	3	30	5.5
10	1	0.1	20	4.47
10	3	0.1	20	4.47
10	5	0.1	20	4.47
30	1	0.05	75	8.66
30	5	0.05	75	8.66
30	10	0.05	75	8.66

As demonstrated in the example plotted in Figures 2.6 and 2.24, the positioning of the precise count y_i within the interval $[j_i, k_i]$ has an impact on the relative shape and positioning of the joint likelihoods for (λ, α). We incorporate this consideration into our simulation experiment by controlling the tendency of y_i to be near j_i or k_i for a given design point. As noted, after generating each y_i we randomly select its position, p_i, within the interval. In those design points for which y_i tends toward j_i, the selection probabilities are weighted more heavily toward lower values of p_i. Similarly, the probabilites are weighted more heavily toward higher values of p_i for design points in which y_i tends toward k_i. For y_i lacking a tendency, the selection probabilities are uniform—all possible values of p_i are equally likely. For those design points in which $w = 1$, we consider $y_i = j_i \ \forall \ i$ versus $y_i = k_i \ \forall \ i$, with y_i allowed to equal j_i or k_i with probability 0.5 each when we prefer no tendency.

We run a *WinBUGS* analysis on the precise data and the interval-censored data for each of the iterations repeated at each design point, as well as on the three data sets based on the alternative methods of dealing with interval-censored data detailed above. Thus, within one iteration we compute posterior estimates from five different methods given the generated data sets. The priors for each Bayesian model are $\pi \sim \text{beta}(1, 1)$ and $r \sim \text{uniform}(0, 1000)$, as discussed in Section 2.5.2. For a single design point we run 100 iterations. Within each iteration we record the posterior means to serve as point estimates, as well as the posterior standard deviations and equal-tailed 95% credible sets. We also obtain maximum likelihood estimates with corresponding standard errors and Wald 95% interval estimates using the Newton-Raphson method described in Section 2.4.1. Because we obtain frequentist estimates, comparisons are made between the frequentist and Bayesian paradigms as well as among factors such as interval width and precise count positioning.

We must use caution in selecting methods of model comparison. Many traditional model selection measures, such as those based on the deviance information criteria, “are comparable only over models with exactly the same observed data” (Spiegelhalter *et al.* 2003). In this simulation we fit five models to each set of data, but each model essentially alters the observed data based on the method employed. The precise data, $\mathbf{y}_n$, changes under each of the precise methods (i.e., original precise data, lower method, mean method, upper method) and the interval-censored data, $\mathbf{d}_n$, further differs from any of the precise count models. Because of this, we focus our comparisons on point and interval estimates for the parameters.

2.7.1 Results for $\lambda = 30$ and $\alpha = 0.05$

In this section we discuss simulation results for the case of $\lambda = 30$ and $w_i = 10 \; \forall \; i$. Results for all cases of $\lambda = 10$ were similar and are summarized in Appendix C. We first summarize the bias in the point estimates. For a given parameter θ, the bias is $\theta - \hat{\theta}$, where θ is the true value from which we generated the data and $\hat{\theta}$ is the point estimate given by the posterior mean (Bayesian approach) or maximum likelihood estimate (frequentist approach). The average bias for a given method is represented by the distance of the means of the point estimates from the true values of the parameters in Figures 2.25 and 2.26.

We also present 95% interval coverage in the figures, by plotting the simulation means of the lower and upper 95% interval endpoints along with shaded boxes. The shaded boxes represent 1 standard deviation above and below average posterior means and interval limits. Coverage is demonstrated by the inclusion of the true parameter value in a design point’s plotted summary interval. From left to right within each panel, the methods plotted are the precise, interval-censored, mean method, lower method, and upper method data.

The simulation summaries on λ in Figure 2.25 look nearly identical for the frequentist and Bayesian approaches. This is expected due to the relatively diffuse priors employed in the Bayesian approach. The mean method performs similarly to

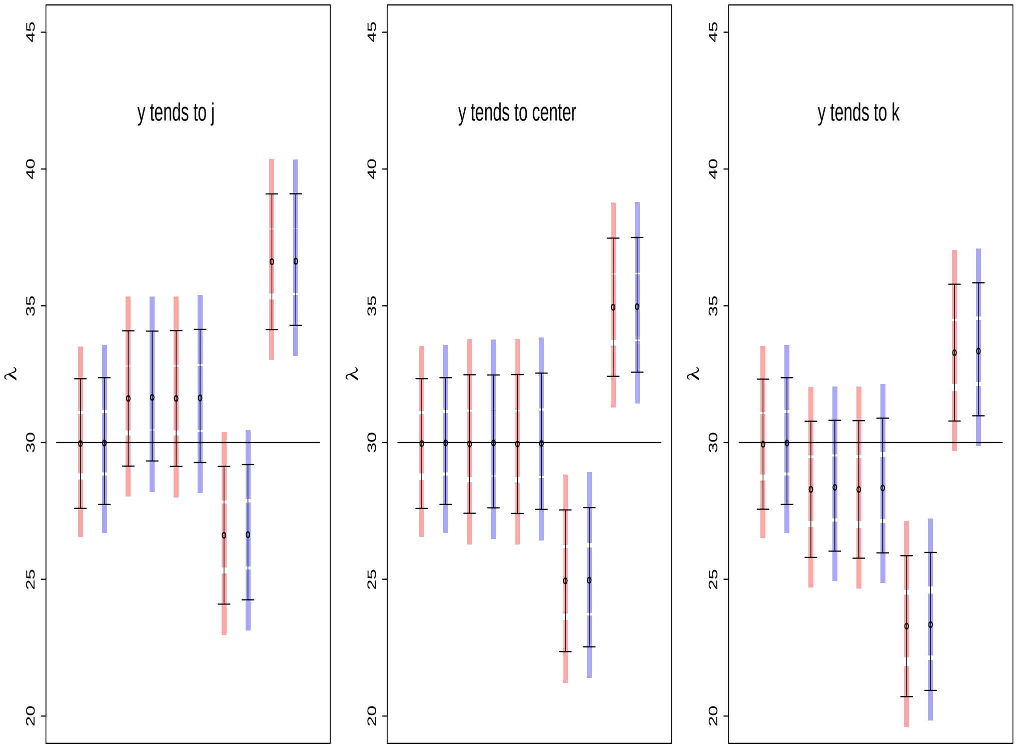

Figure 2.25: Frequentist (red) and Bayesian (blue) simulation results for λ, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 30$ and $\alpha = 0.05$. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

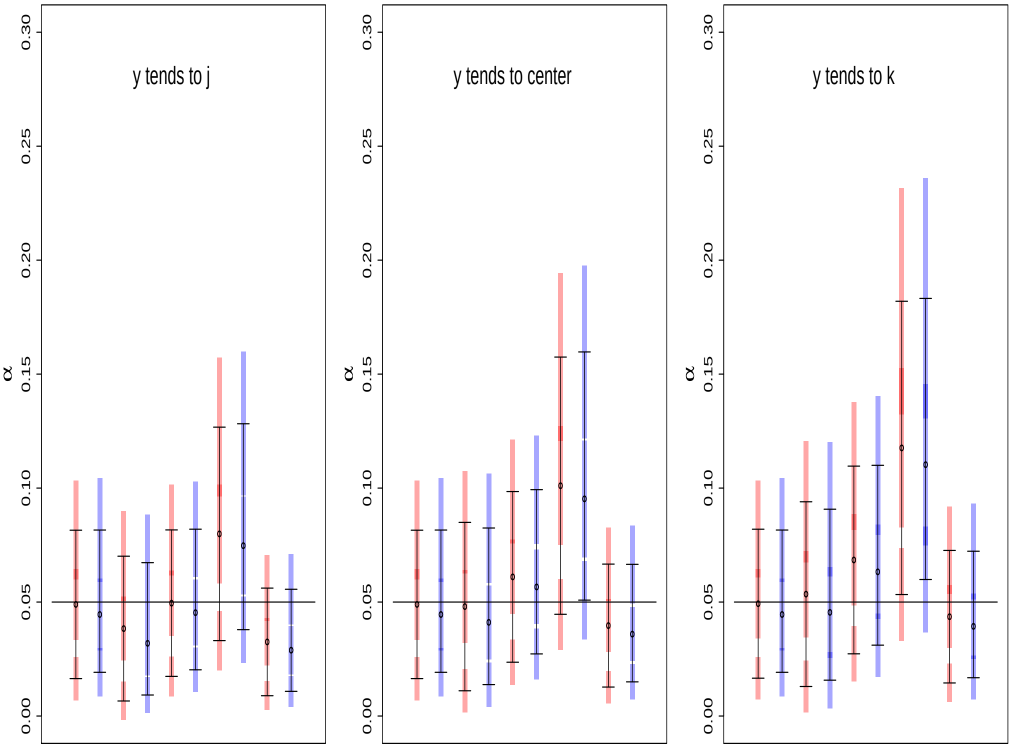

Figure 2.26: Frequentist (red) and Bayesian (blue) simulation results for α, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 30$ and $\alpha = 0.05$. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

the interval-censored method in this example, and these methods are preferable to the alternatives of using $y_i = j_i$ or $y_i = k_i$, even in the scenarios in the first and third figures where the precise data tends toward either of these endpoints. Coverage is poor for the lower and upper methods in all three scenarios.

In the summaries for α plotted in Figure 2.26, simulation mean point estimates vary slightly between the frequentist and Bayesian approaches—the posterior mean is slightly lower than the maximum likelihood estimate in every case. Bayesian credible set widths are also shorter than the frequentist confidence intervals in some cases, but the difference is slight. Again, the interval-censored and mean methods perform similarly, with the lower and upper methods proving to be inferior with greater bias and wider intervals. The interval-censored method exhibits more bias than the mean method when y_i tends toward j_i, the same amount of bias when y_i has no tendency, and less bias when y_i tends to k_i. We summarize interval coverage and width for these simulation results in Tables 2.9 and 2.10. The results confirm the trends seen in the coverage plots.

When we decrease the interval-censored widths in the $\lambda = 30$ case to $w = 5$ and then to $w = 1$, the behavior of the five methods relative to each other remains the same, with the notable difference being the reduction in bias for each width increase across the four methods of dealing with interval-censored data. This is intuitive, as decreasing the interval-censored widths brings the interval-censored data values closer to the original precise data values. The behavior for the five methods is the same in the $\lambda = 10$ case; see Appendix C for a summary of the heavy-censoring case ($w_i = 5 \; \forall \; i$) when $\lambda = 10$.

2.7.2 Results for $\lambda = 3$ and $\alpha = 3$

We see very different behavior from that in Section 2.7.1 when we reduce λ to 3. Results for the case of $\lambda = 3$, $\alpha = 3$, and $w_i = 2 \; \forall \; i$ are plotted in Figures 2.27

Table 2.9. Coverage for λ where true $\lambda = 30$ and IC widths=10

			Freq CI Widths				Bayes CS Widths		
y_i Tend	Method	Est	Avg	SD	Covg	Est	Avg	SD	Covg
j_i	Prec	29.96	4.74	0.48	0.98	31.63	4.63	0.50	0.96
j_i	IC	29.98	4.95	0.51	0.81	26.61	4.75	0.48	0.73
j_i	Mean	31.61	4.97	0.48	0.81	26.63	4.87	0.48	0.74
j_i	Lower	31.65	5.03	0.49	0.22	36.61	4.95	0.49	0.22
j_i	Upper	31.61	4.96	0.48	0.00	36.63	4.81	0.47	0.00
None	Prec	29.96	4.75	0.50	0.98	29.96	4.63	0.50	0.96
None	IC	29.98	5.07	0.54	0.94	24.95	4.86	0.49	0.92
None	Mean	29.95	5.08	0.48	0.95	24.97	4.98	0.48	0.92
None	Lower	29.98	5.18	0.50	0.02	34.95	5.09	0.51	0.03
None	Upper	29.95	5.05	0.48	0.00	34.97	4.93	0.48	0.00
k_i	Prec	29.94	4.76	0.48	0.97	28.34	4.63	0.50	0.96
k_i	IC	29.98	4.98	0.52	0.70	23.29	4.78	0.55	0.68
k_i	Mean	28.29	5.03	0.53	0.72	23.34	4.92	0.55	0.73
k_i	Lower	28.36	5.15	0.59	0.00	33.29	5.04	0.61	0.00
k_i	Upper	28.29	5.00	0.51	0.23	33.34	4.86	0.53	0.20

Table 2.10. Coverage for α where true $\alpha = 0.05$ and IC widths=10

			Freq CI Widths				Bayes CS Widths		
y_i Tend	Method	Est	Avg	SD	Covg	Est	Avg	SD	Covg
j_i	Prec	0.05	0.07	0.01	0.92	0.05	0.06	0.01	0.92
j_i	IC	0.04	0.06	0.01	0.84	0.08	0.06	0.01	0.78
j_i	Mean	0.04	0.06	0.01	0.94	0.07	0.06	0.01	0.93
j_i	Lower	0.03	0.09	0.02	0.90	0.03	0.09	0.02	0.83
j_i	Upper	0.05	0.05	0.01	0.65	0.03	0.04	0.01	0.61
None	Prec	0.05	0.07	0.01	0.92	0.06	0.06	0.01	0.92
None	IC	0.04	0.07	0.01	0.94	0.10	0.07	0.01	0.91
None	Mean	0.05	0.07	0.01	0.97	0.10	0.07	0.01	0.94
None	Lower	0.04	0.11	0.02	0.62	0.04	0.11	0.02	0.48
None	Upper	0.06	0.05	0.01	0.87	0.04	0.05	0.01	0.85
k_i	Prec	0.05	0.07	0.01	0.93	0.06	0.06	0.01	0.92
k_i	IC	0.04	0.08	0.02	0.97	0.12	0.08	0.02	0.94
k_i	Mean	0.05	0.08	0.02	0.94	0.11	0.08	0.02	0.90
k_i	Lower	0.05	0.13	0.03	0.47	0.04	0.12	0.03	0.36
k_i	Upper	0.07	0.06	0.01	0.87	0.04	0.06	0.01	0.84

and 2.28. Again, within each panel the frequentist estimates (red) and Bayesian estimates (blue) are plotted for the five data methods: precise, interval-censored, mean, lower, upper.

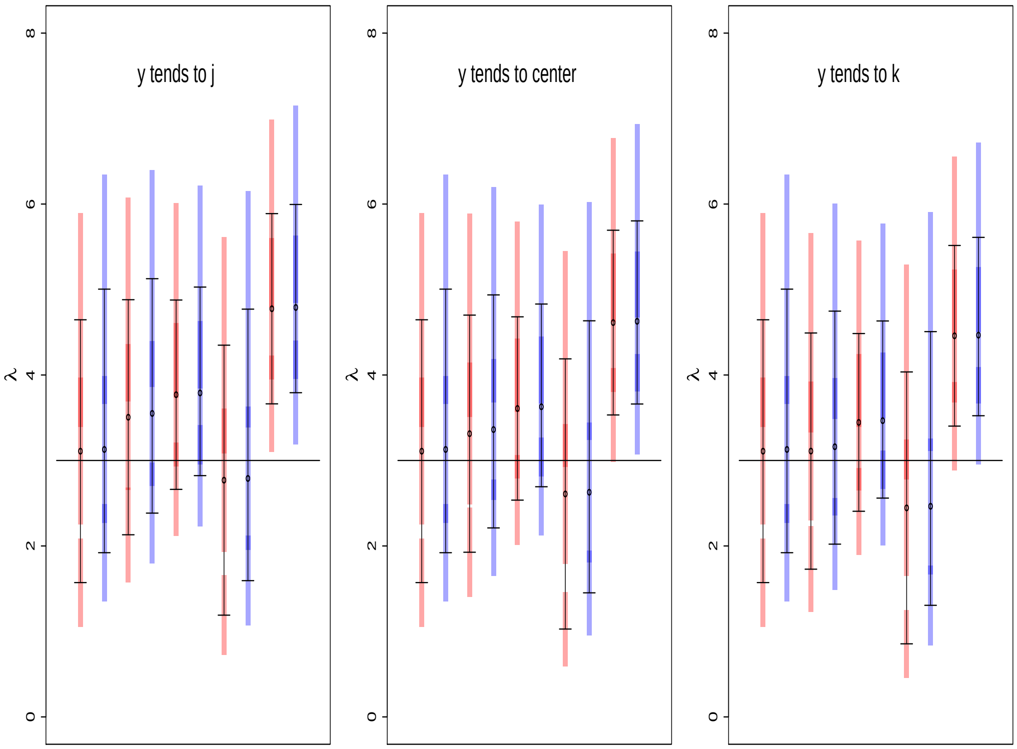

Figure 2.27: Frequentist (red) and Bayesian (blue) simulation results for λ, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 3$ and $\alpha = 3$. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

A key result in the $\lambda = 30$ and $\lambda = 10$ cases was the very similar performance of the interval-censored and mean methods. This $\lambda = 3$ case demonstrates that the two methods do not always correspond. The average bias in estimating λ decreases for the interval-censored method as y_i changes tendency from the left panel to the right panel in Figure 2.27. The bias for the mean method, however, is larger than the bias for the interval-censored method when y_i tends toward j_i and holds steady as y_i changes tendency. Overall, there is not as much differentiation between the behaviors in the three panels as there was for $\lambda = 30$, in Figure 2.25, but this can be attributed to the smaller interval-censored widths. Again, the lower and upper methods have bias and coverage that are inferior to the first three methods, regardless of the tendency of y_i. The interval estimates for the five methods are summarized in Table 2.11.

Table 2.11. Coverage for λ where true $\lambda = 3$ and IC widths=2

			Freq CI Widths				Bayes CS Widths		
y_i Tend	Method	Est	Avg	SD	Covg	Est	Avg	SD	Covg
j_i	Prec	3.11	3.07	0.84	0.93	3.79	3.08	0.85	0.89
j_i	IC	3.13	2.75	0.75	0.88	2.77	2.74	0.76	0.80
j_i	Mean	3.51	2.22	0.61	0.71	2.79	2.21	0.61	0.62
j_i	Lower	3.55	3.16	0.91	0.85	4.78	3.18	0.92	0.90
j_i	Upper	3.77	2.23	0.55	0.09	4.79	2.20	0.57	0.09
None	Prec	3.11	3.07	0.84	0.93	3.63	3.08	0.85	0.89
None	IC	3.13	2.77	0.79	0.89	2.61	2.73	0.78	0.83
None	Mean	3.31	2.14	0.60	0.73	2.63	2.14	0.60	0.70
None	Lower	3.36	3.16	0.95	0.79	4.61	3.18	0.96	0.89
None	Upper	3.61	2.16	0.54	0.20	4.63	2.14	0.56	0.11
k_i	Prec	3.11	3.07	0.84	0.93	3.47	3.08	0.85	0.89
k_i	IC	3.13	2.76	0.77	0.88	2.45	2.73	0.79	0.85
k_i	Mean	3.11	2.08	0.59	0.78	2.46	2.07	0.60	0.74
k_i	Lower	3.16	3.18	0.97	0.78	4.46	3.20	0.99	0.87
k_i	Upper	3.45	2.11	0.54	0.21	4.47	2.09	0.55	0.20

There is high differentiation among the five data methods in the results for α plotted in Figure 2.28. The precise method yields results that are the least biased, followed by the interval-censored method in the cases where y_i has no tendency or tends toward k_i. When y_i tends toward j_i, the lower method actually presents less bias than the interval-censored method, but at a cost of much higher interval-estimate widths. The mean method has greater precision (represented by shorter interval-esimate widths) across the panels, but at a value that is highly biased from the true α. Thus, in the case of interval-censored data, the interval-censored method came closer to estimating the true α than the other three methods. The interval estimates of α for the five methods are summarized in Table 2.12.

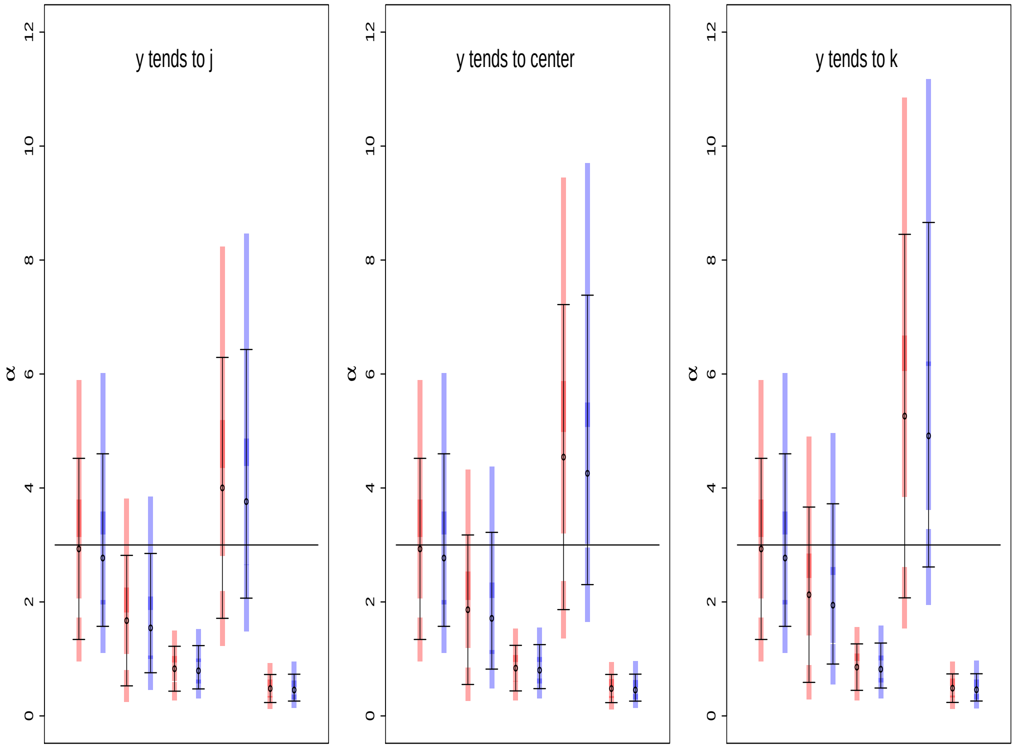

Figure 2.28: Frequentist (red) and Bayesian (blue) simulation results for α, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 3$ and $\alpha = 3$. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

2.7.3 Discussion

In this simulation study, we began with precise negative binomial data and applied interval-censoring to those counts, then compared four different methods of handling the interval-censored data in terms of ability to estimate the true values of λ and α. Using the mean, lower, or upper methods might be a resarcher's first instinct in handling interval-censored data, as a means of forcing precise counts and proceeding with a typical negative binomial model. For example, the CDC recommends calculating the "unhealthy days index" introduced in Chapter 1 by taking the upper endpoints of the intervals. We introduced the interval-censored model as an alternative. We compared the average bias of the point estimates across simulation reps, as well as the average width and coverage of the interval estimates.

Table 2.12. Coverage for α where true $\alpha = 3$ and IC widths=2

			Freq CI Widths				Bayes CS Widths		
y_i Tend	Method	Est	Avg	SD	Covg	Est	Avg	SD	Covg
j_i	Prec	2.93	3.18	1.04	0.84	0.79	3.03	0.98	0.83
j_i	IC	2.77	2.29	0.90	0.38	4.00	2.09	0.73	0.42
j_i	Mean	1.67	0.79	0.12	0.00	3.76	0.76	0.12	0.00
j_i	Lower	1.54	4.58	1.53	0.99	0.48	4.36	1.49	0.94
j_i	Upper	0.83	0.49	0.09	0.00	0.46	0.47	0.10	0.00
None	Prec	2.93	3.18	1.04	0.84	0.80	3.03	0.98	0.83
None	IC	2.77	2.62	1.01	0.55	4.54	2.40	0.86	0.54
None	Mean	1.86	0.80	0.12	0.00	4.26	0.77	0.13	0.00
None	Lower	1.71	5.35	1.82	0.99	0.48	5.08	1.71	0.87
None	Upper	0.84	0.49	0.10	0.00	0.45	0.47	0.11	0.00
k_i	Prec	2.93	3.18	1.04	0.84	0.82	3.03	0.98	0.83
k_i	IC	2.77	3.08	1.10	0.73	5.26	2.81	0.95	0.74
k_i	Mean	2.13	0.82	0.12	0.00	4.91	0.79	0.13	0.00
k_i	Lower	1.94	6.38	2.02	0.92	0.49	6.05	1.92	0.74
k_i	Upper	0.86	0.50	0.10	0.00	0.46	0.48	0.11	0.00

We began by setting the true value of $\lambda = 30$. Initial findings revealed that for large λ, if α is also large so as to cause extreme variability in the generated data, little differentiation can be found among the five methods, as all interval estimates are very wide. This motivated our choice of $\alpha = 0.05$ when $\lambda = 30$ and $\alpha = 0.1$ when $\lambda = 10$. For these two cases, we found that the interval-censored and mean methods had very similar performance in the simulation results. Their average bias was exactly the same in the results for λ, and this bias was positive for y_i tending toward j_i, zero for y_i lacking tendency, and negative for y_i tending toward k_i. The lower and upper methods were inferior for estimating λ in this case, with large biases and interval coverages near zero. In the estimates for α, the interval-censored method slightly reduced bias relative to the mean method as y_i shifted in tendency from j_i to k_i. Overall, the simulations under this case demonstrated that the interval-censored or mean methods are superior to the lower or upper methods, even when y_i tends toward the lower or upper endpoints, at least in the scenarios studied.

These results might cause a researcher to favor the mean method rather than the more complicated interval-censored model; however, our results in Section 2.7.2 demonstrated that the interval-censored and mean methods do not always agree in their results. Here we had a smaller λ of 3, with $w_i = 2 \ \forall \ i$, representing heavy censoring relative to the data. The bias in the λ estimates was smallest for the interval-censored and lower methods, but the lower method consistently presented wider interval estimates, so the interval-censored method would be preferred. Similarly, in the estimates for α, The interval-censored and lower methods revealed the least amount of bias. The lower method exhibited better coverage of the true α than the interval-censored method for all tendencies of y_i, but the average width was also several units wider under each tendency. Overall, when taking estimation of both λ and α into account, we would prefer the interval-censored method here.

These findings demonstrated that, for the cases under our consideration, the interval-censored method performs at least as well as (and sometimes better than) the other methods for handling interval-censored data. In future work, we will investigate other features not considered in this study. For example, in each data set for a given design point we fixed w_i at one value $\forall \ i$. A natural extension would be to allow w_i to take on various values. We will also investigate varying sample sizes and percentage of the data that are censored.

CHAPTER THREE

Interval-Censored Negative Binomial Regression

We now extend the interval-censored negative binomial model detailed in Chapter 2 to incorporate regression. We utilize the generalized linear model (GLM) framework. Suppose we have a collection of n independent observations $Y_1, \ldots, Y_n$ from negative binomial distributions with dispersion parameters α_i and mean parameters $\lambda_i|\mathbf{x}_i, \boldsymbol{\beta}$ which are dependent on vectors of covariates, $\mathbf{x}_i$, and regression coefficients, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, such that Y_i is a positive integer. We have

$$Y_i \sim NB(\lambda_i, \alpha_i),$$

and

$$g(\lambda_i) = \log(\lambda_i) = \mathbf{x}_i^{'}\boldsymbol{\beta}$$

for $i = 1, 2, \ldots, n$, where α_i and λ_i are the parameters for each individual i. The function g is a one-to-one differentiable link function, such that

$$\begin{aligned} g[E(Y_i|\mathbf{x}_i, \boldsymbol{\beta})] &= g(\lambda_i|\mathbf{x}_i, \boldsymbol{\beta}) \\ &\equiv g(\lambda_i) \\ &= \mathbf{x}_i^{'}\boldsymbol{\beta}. \end{aligned}$$

Then $\lambda_i = g^{-1}(\mathbf{x}_i^{'}\boldsymbol{\beta})$. We choose $g(\lambda_i) = \log(\lambda_i)$ which is the usual link function when modeling overdispersed Poisson data. Hilbe (2011, p. 4) notes that the canonical link is $-\log((1/\alpha\lambda) + 1)$, but the log link relates the negative binomial regression model to the Poisson model more directly. Then we have $\lambda_i = \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})$. If we observe $(Y_1, \ldots, Y_n) = (y_1, \ldots, y_n) \equiv \mathbf{y}$, the likelihood function for $\boldsymbol{\beta}$ given $\mathbf{y}$ and $\boldsymbol{\alpha}$, the vector of dispersion parameters, is

$$L(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \frac{1}{\alpha_i})}{y_i!\Gamma(\frac{1}{\alpha_i})} \Big(1 + \alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})\Big)^{1/\alpha_i} \left(\frac{\alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})}{1 + \alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})}\right)^{y_i}. \quad (3.1)$$

Now, as in Chapter 2, suppose that we do not observe all counts precisely. Instead, for some subset of the observations, the ith count is known only to fall within the interval $[j_i, k_i]$ with probability one. That is, $Y_i \in [j_i, j_i + 1, \ldots, k_i]$ with probability one, for all i. Some observations may still be precise, in which case $j_i = k_i$. Suppose then that we observe $[j_i, k_i]$, $i = 1, \ldots, n$, where $j_i \leq k_i \ \forall \ i$. Denote by $\mathbf{d}_n$ the collection of these intervals. That is, $\mathbf{d}_n = \{[j_i, k_i]\}_{i=1}^{n}$. Then the likelihood for $\boldsymbol{\beta}$ given $\mathbf{d}_n$ is

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{d}_n, \boldsymbol{\alpha}) = &\prod_{i=1}^{n} \sum_{y_i=j_i}^{k_i} \frac{\Gamma(y_i + \frac{1}{\alpha_i})}{y_i!\Gamma(\frac{1}{\alpha_i})} \left(1 + \alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})\right)^{1/\alpha_i} \\ &\times \left(\frac{\alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})}{1 + \alpha_i \exp(\mathbf{x}_i^{'}\boldsymbol{\beta})}\right)^{y_i}. \end{aligned} \tag{3.2}$$

3.1 General Social Survey Example

To illustrate the use of (3.2) and the corresponding frequentist and Bayesian analyses, we present an example in which we use the negative binomial to model overdispersed Poisson data. Ntzoufras (2009, pp. 284–286) analyzes data from $n = 550$ respondents to the 1990 U.S. General Social Survey (GSS). In one particular question on the survey, respondents were asked to report their total number of sexual intercourses in the previous month. This data set might be of interest, for example, to epidemiologists studying sexually transmitted diseases. Gender, along with other covariates, was recorded for each respondent. Ntzoufras found that the sample variances of the sexual intercourse count were much higher than the sample means for both men and women; thus he introduced a negative binomial regression model and compared this to the Poisson model. The sample mean and variance were given by 5.9 and 54.8 for men, and 4.3 and 34.3 for women. Precise counts were given in this data set; however, we might imagine a scenario in which the researcher does not require the respondent to give a precise answer on data of this nature. Then each respondent may give a range, e.g., “I know it was between 3 and 5,” and the researcher would be presented with an interval-censored count data set. This is the scenario we use to illustrate many of the methods developed below, using gender as the single binary covariate.

Denote by G_i an indicator variable on gender, with $G_i = 1$ indicating females. Consider the model

$$Y_i \sim NB(\alpha_i^*, \lambda_i^*)$$

and

$$\log(\lambda_i^*) = \beta_1 + \beta_2 G_i$$

where α_i^* and λ_i^* are the parameters for each individual i ($i = 1, \ldots, 550$) in the original dataset. In what follows, we will analyze the original data and then generate a smaller working data set based on the parameter values estimated from the original data.

Let λ_1 and λ_2 be the expected counts for males and females, respectively; that is,

$$\lambda_1 = e^{\beta_1},\ \lambda_2 = e^{\beta_1+\beta_2}.$$

Further, let α_1 and α_2 represent the dispersion parameters for males and females, respectively, with $r_1 = 1/\alpha_1$ and $r_2 = 1/\alpha_2$. We are primarily interested in λ_1 and λ_2 for their ease of interpretation. That is, we are comparing the expected number of sexual intercourses reported for males versus females. We can also examine β_2 and its interval estimates in order to study the difference in reported number of sexual intercourses for males versus females.

3.2 Frequentist Development

Just as we used iterative methods to obtain MLEs for λ and α in the non-regression context, we now employ Newton-Raphson techniques to find estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the regression scenario. The negative binomial log-likelihood for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ given precise data $\mathbf{y} = (y_1, \ldots, y_n)$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}|\mathbf{y}) \equiv \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha_i \exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\alpha_i \exp(\mathbf{x}_i'\boldsymbol{\beta})}\right) - \frac{1}{\alpha_i}\ln(1+\alpha_i \exp(\mathbf{x}_i'\boldsymbol{\beta})) + \ln\Gamma\left(y_i + \frac{1}{\alpha_i}\right) \\ - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha_i}\right). \qquad (3.3)\end{aligned}$$

The Newton-Raphson method updates the vector of parameters, $\boldsymbol{\beta}_r$, involved in the likelihood (3.3) through the process described in Section 2.4, where the necessary partial derivatives in $\mathbf{U}$ and $\mathbf{H}$ are with respect to the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ rather than α and λ. For more detail regarding such procedures see, for example, Gentle (2009), p. 266ff or Khuri (2003), p. 331.

In our example with a single binary covariate we also derive point estimates for λ_1 and λ_2 and interval estimates using the delta method, based on the relationships (Ntzoufras 2009, p. 285)

$$\lambda_1 = g_1(\beta_1) \equiv \exp(\beta_1) \qquad (3.4)$$

and

$$\lambda_2 = g_2(\beta_1, \beta_2) \equiv \exp(\beta_1 + \beta_2). \qquad (3.5)$$

To implement the Newton-Raphson method for estimates and standard errors we use the `ml.nb2` function from the *R* package `COUNT` (Hilbe *et al.* 2012). This package was created to accompany Hilbe's (2011) text (pp. xv–xvi). The function `ml.nb2` is used to find maximum likelihood estimates using negative binomial data. It assumes the more common parameterization (2.2). The model regressing the count response on a linear combination of covariates must be specified, and output consists of a table of parameter estimates, standard errors, and confidence intervals. By default, the function uses a general purpose optimization procedure based on the method of Nelder and Mead (1965). We alter the function's algorithm to instead employ Newton-Raphson optimization for our parameterization using `maxNR` in place of the standard `optim`. We also alter the code to include calculation of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ by

(3.4) and (3.5) as well as the standard errors of these estimates through the function `deltamethod`.

The maximum likelihood estimates and corresponding 95% Wald confidence intervals from the original data analyzed by Ntzoufras (all precise counts) are given in Table 3.1.

Table 3.1. Frequentist Estimates Based on GSS Data

Param.	Estimate	SE	LCL	UCL
β_1	1.77	0.08	1.60	1.93
β_2	-0.31	0.13	-0.56	-0.06
λ_1	5.86	0.49	4.90	6.81
λ_2	4.30	0.40	3.50	5.09
α_1	1.49	0.17	1.16	1.83
α_2	2.52	0.26	2.00	3.03

Because the 95% confidence interval for $\beta_2 = (-0.56, -0.06)$ does not contain zero, we can conclude that women report a significantly lower number of sexual intercourses for the past month than men. Note that the estimates for λ_1 and λ_2 correspond to the sample means of the original data. In Section 3.4.1, we show that the Wald approximation is sometimes not appropriate.

3.3 Bayesian Development

In the Bayesian approach to this problem, we must define a prior structure for the regression parameters as well as the dispersion parameter in the model given by (3.1). Ntzoufras parameterizes (3.1) in terms of $\pi_i = 1/(1+\alpha_i\lambda_i)$ and $r_i = 1/\alpha_i$:

$$L(\pi_i, r_i|\{y_i\}) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \frac{1}{r_i})}{y_i!\Gamma(r_i)} \pi_i^{r_i}(1-\pi_i)^{y_i}, \tag{3.6}$$

where $\pi_i = (\pi_1, \pi_2)$, $\pi_i = r_i/(r_i + \lambda_i)$, and $\log(\lambda_i) = \beta_1 + \beta_2 G_i$. Ntzoufras provides *WinBUGS* code for his precise GSS data. We implemented that code using a burn-in of 1,000, thinning rate of 20 iterations followed by 10,000 updates. (Our version of his code is available upon request.) Our results using this model are shown in Table 3.2.

Table 3.2. Ntzoufras (2009) *WinBUGS* Analysis

Param.	Post. Mean	SD	Post. Interval
β_1	1.77	0.08	(1.61, 1.94)
β_2	-0.31	0.13	$(-0.55, -0.06)$
λ_1	5.89	0.50	(5.00, 6.93)
λ_2	4.33	0.41	(3.60, 5.21)
α_1	1.52	0.17	(1.20, 1.88)
α_2	2.54	0.27	(2.06, 3.12)

These estimates are very similar to those we obtained via frequentist methods in Section 3.2, as expected given the large sample size and our use of relatively diffuse priors.

3.4 An Example

In this section we present an example based on a relatively small simulated data set in order to contrast the methods outlined to this point. We generate new precise count data based on the parameter estimates from the GSS data described in Section 3.1. The percentage of female respondents in the original data was 56%; for simplicity we will generate an equal number of males and females in our small sample data set. We generate this data set using $\lambda_1 = 6$, $\lambda_2 = 4$, $\alpha_1 = 1.5$ and $\alpha_2 = 2.5$.

We begin with data of sample size $n = 50$ (25 males and 25 females). The sample means and variances for this data are given in Table 3.3.

In what follows, we refer to this as the "small GSS data set." First we find frequentist and Bayesian estimates for the generated precise data. For the frequentist approach we employ our modified `ml.nb2` function as described in Section 3.2. The Newton-Raphson optimization process requires initial values for the parameters; we

Table 3.3. Small GSS Data Set Summary

Gender	True Mean	Sample Mean	True Var	Sample Var
Males	6	6.48	60	63.34
Females	4	3.56	44	31.51

initialize α_1 and α_2 at 0.5, β_1 at -1, and β_2 at zero. This is the convention employed in the original `ml.nb2` *R* function. After the specified tolerance is reached, the function is used to find the frequentist standard errors using the square root of the diagonal elements of the negative inverse of the final Hessian matrix. These standard errors are used to calculate Wald 95% confidence intervals which, for a given MLE $\hat{\theta}$ and its standard error, $SE_{\hat{\theta}}$, are given by

$$\hat{\theta} \pm 1.96 \times SE_{\hat{\theta}}.$$

The MLEs and corresponding standard errors for λ_1 and λ_2 are calculated after the Newton-Raphson iterations converge, using the delta method as described in Section 3.2.

Our Bayesian analysis employs Ntzoufras's prior choice of $\beta_j \sim N(0, 1000)$ and $r_j \sim \text{gamma}(0.001, 0.001)$ for $j = 1, 2$. We use two chains with a burn-in of 1,000, followed by 10,000 updates with a thinning rate of 10. The parameters λ_1 and λ_2 are defined within the *WinBUGs* code and monitored through the updates, so this differs from the frequentist approach in that there is no need to estimate standard errors for λ_1 and λ_2 using the delta method after the analysis. Diagnostic techniques including the autocorrelation plots indicate convergence for these settings. We compare the results from the frequentist and Bayesian analyses in Table 3.4.

Though the point estimates of the regression coefficients are quite similar for the frequentist and Bayesian methods, there is slightly more differentiation in the estimates for the α's and λ's. Additionally, the 95% Bayesian credible sets are wider than the 95% frequentist confidence intervals.

Table 3.4. Frequentist and Bayesian Estimates for Small GSS Data Set

Param	Truth	MLE	Post. Mean	SE	Post. SD	CI	CS
β_1	1.79	1.87	1.89	0.23	0.25	(1.41,2.33)	(1.43,2.42)
β_2	-0.41	-0.60	-0.57	0.40	0.44	$(-1.39, 0.19)$	$(-1.40, 0.34)$
λ_1	6.00	6.48	6.86	1.52	1.82	(3.50,9.45)	(4.17,11.27)
λ_2	4.00	3.56	4.04	1.17	1.81	(1.27, 5.85)	(1.95,8.17)
α_1	1.50	1.22	1.37	0.40	0.47	(0.44,2.00)	(0.68,2.49)
α_2	2.50	2.40	2.75	0.85	1.04	(0.74, 4.06)	(1.30,5.27)

Next we create a data set with relatively mild interval centering from the small GSS data set by setting $w_i = 1 \ \forall \ i$ and randomly selecting whether $y_i = j_i$ or $y_i = k_i$ for each set. For example, the first precise point in the small GSS data set is $y_1 = 19$. We generate the interval $[j_1, k_1] = [19, 20]$ for our first interval-censored observation.

Table 3.5. Bayesian Estimates for Precise Small GSS Data Set and IC with $w_i = 1 \ \forall \ i$

Param	Truth	Prec Mean	IC Mean	Prec SD	IC SD	Prec CS	IC CS
β_1	1.79	1.89	1.92	0.25	0.24	(1.43,2.42)	(1.48, 2.41)
β_2	-0.41	-0.57	-0.59	0.44	0.46	$(-1.40, 0.34)$	$(-1.45, 0.36)$
λ_1	6.00	6.86	7.02	1.82	1.76	(4.17,11.27)	(4.37, 11.16)
λ_2	4.00	4.04	4.09	1.81	2.01	(1.95,8.17)	(1.87,8.67)
α_1	1.50	1.37	1.24	0.47	0.45	(0.68,2.49)	(0.60,2.32)
α_2	2.50	2.75	3.19	1.04	1.47	(1.30,5.27)	(1.30,6.83)

The second precise point is $y_2 = 5$ and we generate $[j_2, k_2] = [4, 5]$. The precise Bayesian estimates from Table 3.4 are given together with the estimates resulting from interval-censored data, in Table 3.5.

As we might expect, there is not much of a difference in the parameter estimates when interval censoring widths are limited to one unit. The main difference in this particular example appears in the estimates for α_2, the dispersion parameter for females. The posterior mean of α_2 from the interval-censored data is higher and the interval estimate is wider.

To demonstrate the effect of more extreme interval censoring for this example, we set $w_i = 5 \ \forall \ i$ and run a new *WinBUGS* analysis. This may not be a realistic case of interest as the interval widths surpass the expected count for females, but it demonstrates wider intervals in the data. The results for the precise and interval-censored data are given in Table 3.6.

Table 3.6. Bayesian Estimates for Precise Small GSS Data Set and IC with $w_i = 5 \;\forall\; i$

Param	Truth	Prec Mean	IC Mean	Prec SD	IC SD	Prec CS	IC CS
β_1	1.79	1.89	1.92	0.25	0.24	(1.43,2.42)	(1.46, 2.39)
β_2	-0.41	-0.57	-0.47	0.44	0.45	$(-1.40, 0.34)$	$(-1.34, 0.44)$
λ_1	6.00	6.86	7.00	1.82	1.72	(4.17,11.27)	(4.31, 10.96)
λ_2	4.00	4.04	4.63	1.81	2.22	(1.95,8.17)	(2.01,9.46)
α_1	1.50	1.37	1.15	0.47	0.57	(0.68,2.49)	(0.44,2.55)
α_2	2.50	2.75	3.00	1.04	2.31	(1.30,5.27)	(0.67,9.08)

The posterior standard deviations for λ_2, α_1, and α_2 are higher in the interval-censored case than in the precise case, and the posterior means for these estimates are farther from the truth than in the precise case.

3.4.1 Appropriateness of the Wald Approximation

We have used Wald intervals in constructing our frequentist interval estimates thus far. The Wald formula relies on the approximate asymptotic normality of the estimator. The requisite regularity of the likelihood (Pawitan, 2001, p. 241) may not obtain in some problems and, as we have seen in Chapter 2, can fail in our interval censoring context. Similar to our exploration in Chapter 2, we will analyze the distributions of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\alpha}}$ to determine whether interval estimates based on an alternative approximation may be needed in our particular data scenario. To this end, we will use the small GSS data set in Section 3.4 and use an interval-censored data set with $w_i = 5 \;\forall\; i$, in order to demonstrate the effects of more severe censoring on the sampling distributions of the estimates.

To obtain approximate sampling distributions we generated 1000 precise data sets and corresponding interval-censored data sets with $w_i = 5 \;\forall\; i$. We used the same parameter values employed to generate our small GSS data set of size $n = 50$, with $\lambda_1 = 6$, $\lambda_2 = 4$, $\alpha_1 = 1.5$ and $\alpha_2 = 2.5$. We find $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\alpha}}$ for the precise and interval-censored data in each replication based on the Newton-Raphson iterative scheme, then plot histograms and normal quantiles of the generated MLEs. The distributions of the estimates for β_1 and β_2 are plotted in Figure 3.1. The empirical sampling distributions for $\hat{\beta}_1$ and $\hat{\beta}_2$ appear to be roughly bell-shaped with no clear issues in the normal quantile plots. We plot distributions of the estimates for λ_1 and λ_2 in Figure 3.2. There is some skewness in the distributions of the estimators, especially in $\hat{\lambda}_2$, indicating that the normal approximation may not be adequate. In Figure 3.3 we plot the histograms and normal quantiles for $\hat{\alpha}_1$ and $\hat{\alpha}_2$. The distributions for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are highly skewed, with large deviations from the straight line in the QQ plot. This indicates that the quadratic approximation would not be appropriate for these distributions, and thus the Wald formula for interval estimates of $\hat{\boldsymbol{\alpha}}$ is not the optimal choice.

In Figure 3.4 we plot the quadratic approximations in blue for each MLE for β_1 and β_2, superimposed over the histograms of the 1000 generated MLEs. The quadratic approximation seems to provide an adequate fit for the sampling distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ in both the precise and interval-censored cases, indicating that the Wald intervals for these parameters may be suitable. Next we plot the quadratic approximations for the sampling distributions of $\hat{\boldsymbol{\lambda}}$ in Figure 3.5.

The approximations in the precise case seem to fit fairly well, although the right-tail behavior of the approximations deviates from the empirical distributions. The approximations in the interval-censored case are not as accurate, with peaks to the left of the empirical peaks and poor tail behavior. In Figure 3.6 we plot the approximations for $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Similar to the behavior in Figure 3.5, the quadratic approximations provide better fits for the sampling distributions of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in the precise case than in the interval-censored case. The peaks are inaccurate and there is improper tail behavior in the interval-censored case for $\hat{\alpha}_1$.

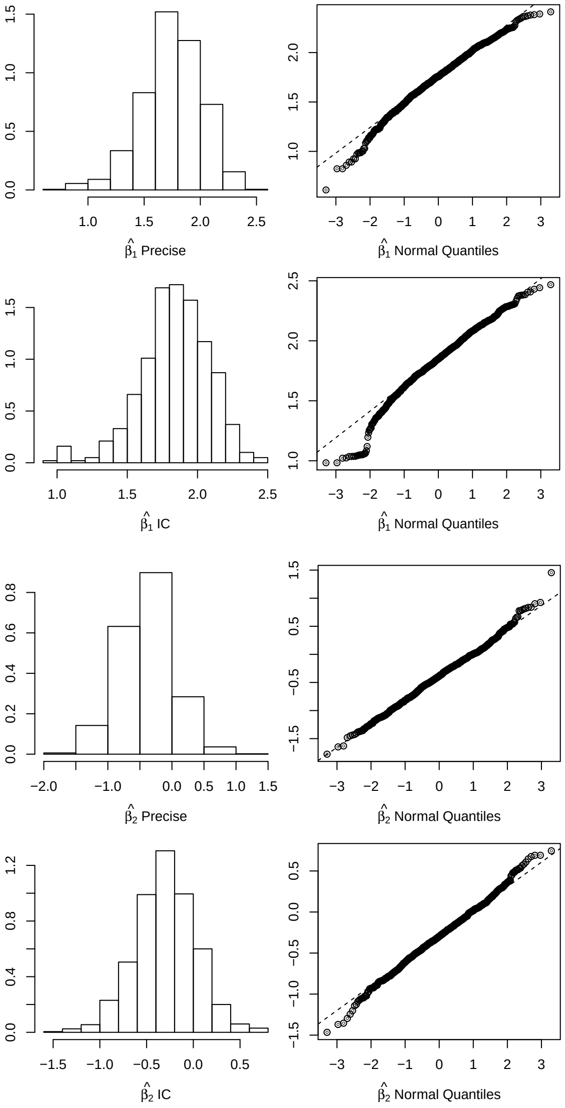

Figure 3.1: Histograms of the empirical sampling distributions based on 1000 datasets of size $n = 50$ of the MLEs for β_1 based on precise data (top left panel), β_1 based on interval-censored data (second left panel), β_2 based on precise data (third left panel), and β_2 based on interval-censored data (bottom left panel), with corresponding normal quantile plots (right panels).

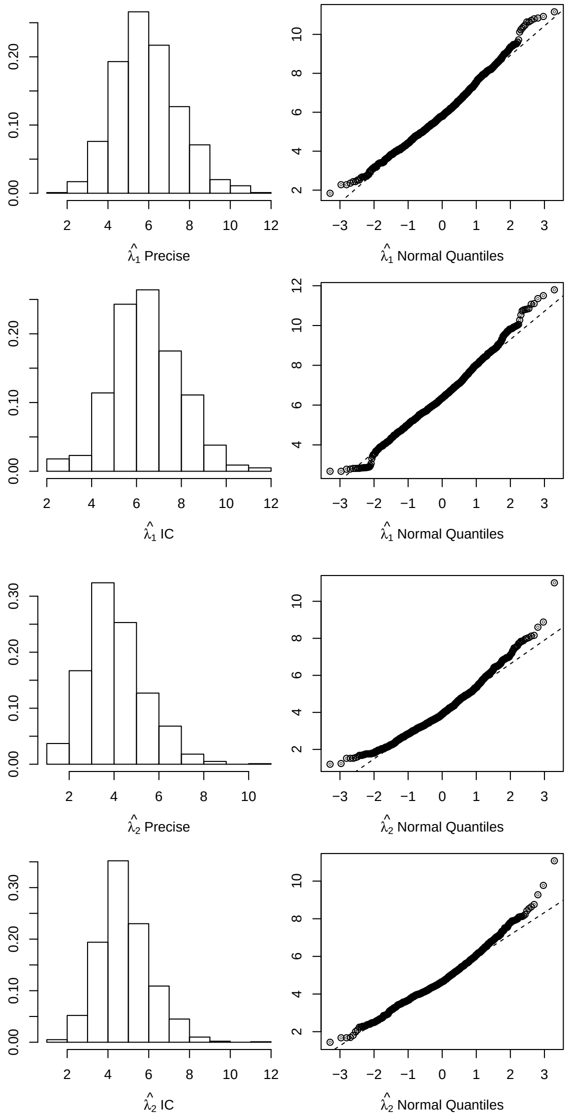

Figure 3.2: Histograms of the empirical sampling distributions based on 1000 datasets of size $n = 50$ of the MLEs for λ_1 based on precise data (top left panel), λ_1 based on interval-censored data (second left panel), λ_2 based on precise data (third left panel), and λ_2 based on interval-censored data (bottom left panel), with corresponding normal quantile plots (right panels).

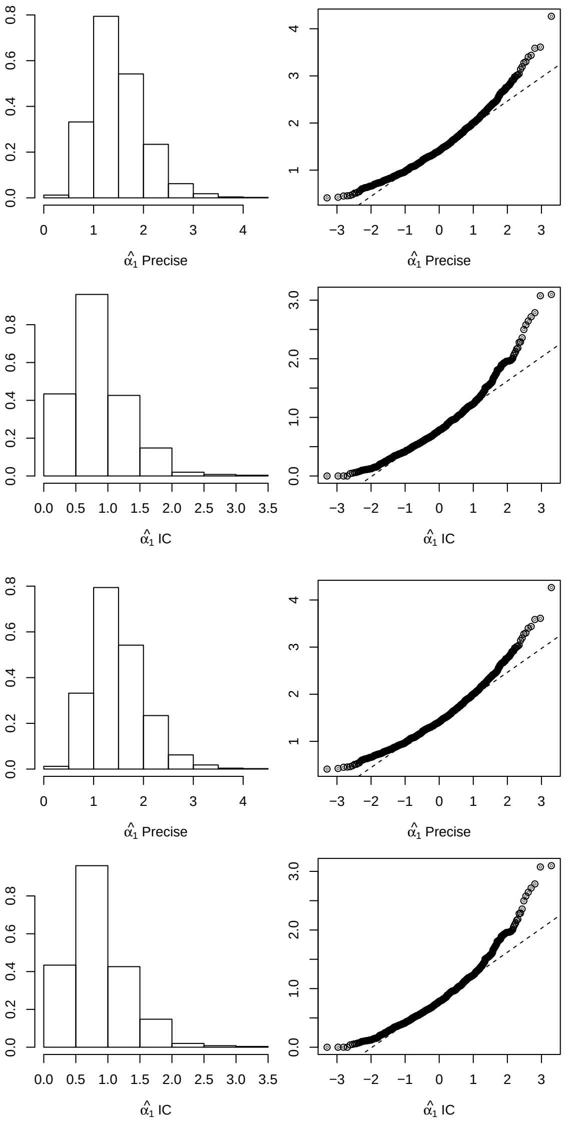

Figure 3.3: Histograms of the empirical sampling distributions based on 1000 datasets of size $n = 50$ of the MLEs for α_1 based on precise data (top left panel), α_1 based on interval-censored data (second left panel), α_2 based on precise data (third left panel), and α_2 based on interval-censored data (bottom left panel), with corresponding normal quantile plots (right panels).

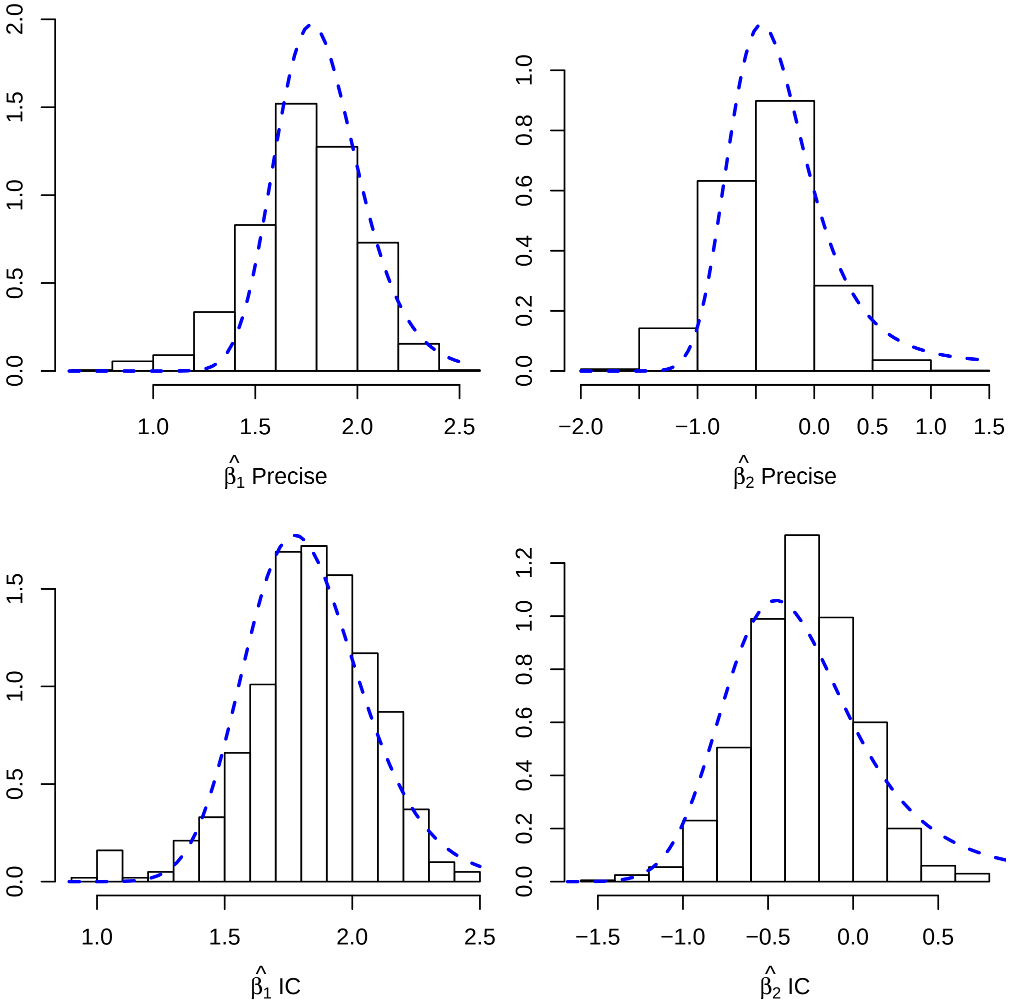

Figure 3.4: Histograms of maximum likelihood estimators based on 1000 replications of samples of size $n = 50$ with normal theory approximations to the corresponding densities superimposed (blue dashed line). The left set of panels are for estimators $\hat{\beta}_1$; the right for estimators $\hat{\beta}_2$. The top row of panels is for precise data; the bottom row for interval-censored data with $w = 5$.

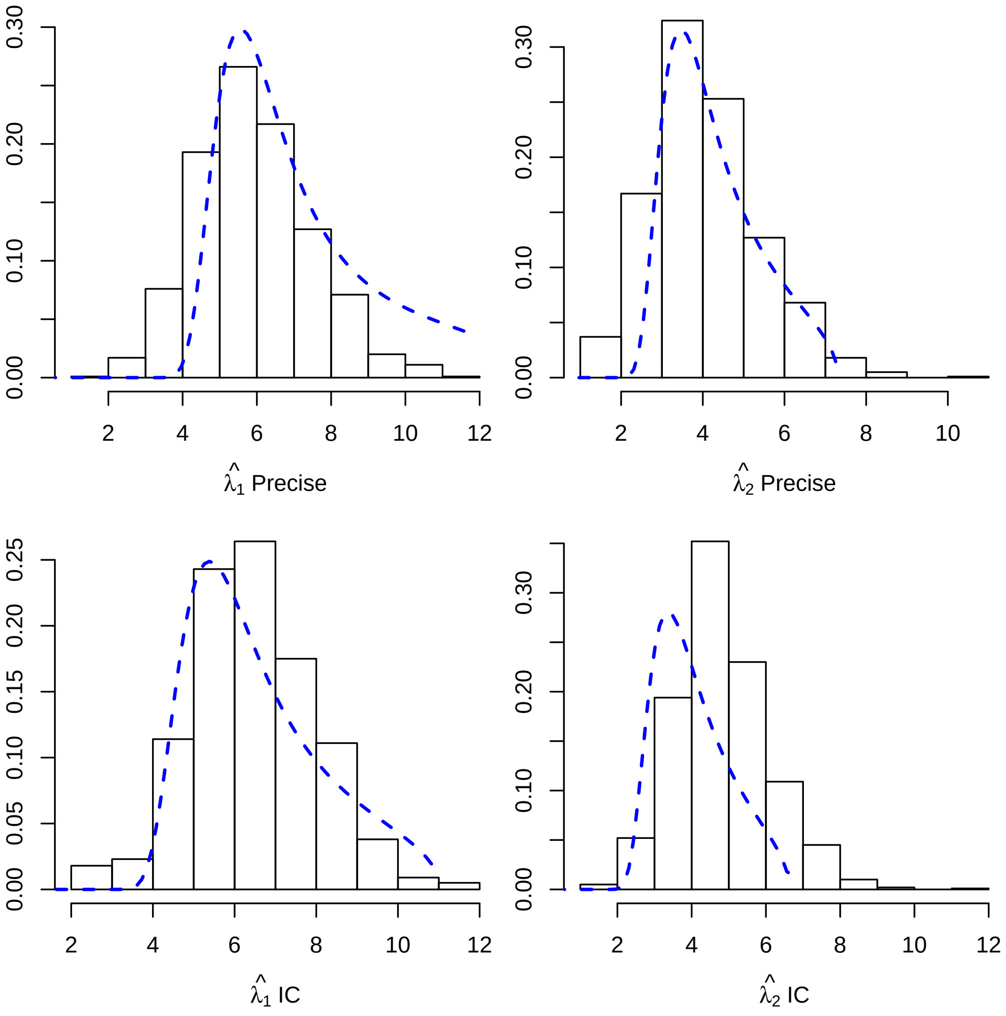

Figure 3.5: Histograms of maximum likelihood estimators based on 1000 replications of samples of size $n = 50$ with normal theory approximations to the corresponding densities superimposed (blue dashed line). The left set of panels are for estimators $\hat{\lambda}_1$; the right for estimators $\hat{\lambda}_2$. The top row of panels is for precise data; the bottom row for interval-censored data with $w = 5$.

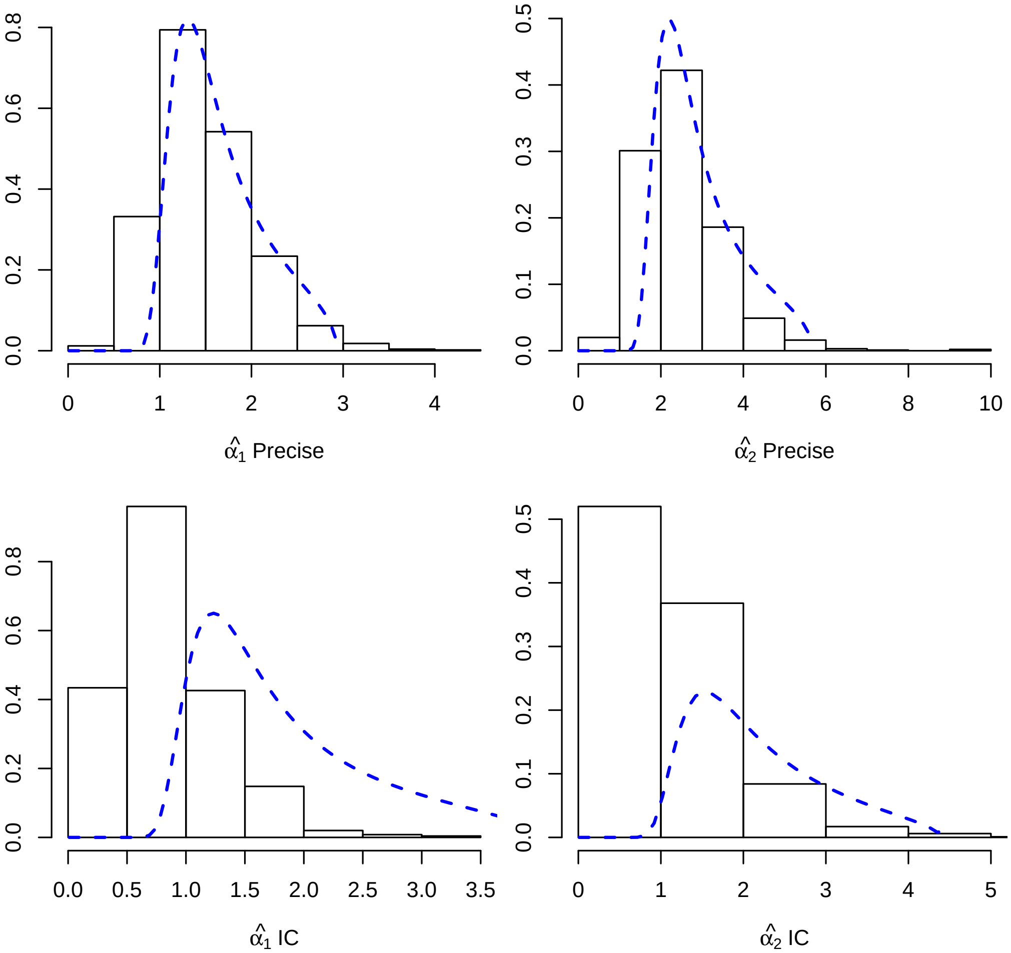

Figure 3.6: Histograms of maximum likelihood estimators based on 1000 replications of samples of size $n = 50$ with normal theory approximations to the corresponding densities superimposed (blue dashed line). The left set of panels are for estimators $\hat{\alpha}_1$; the right for estimators $\hat{\alpha}_2$. The top row of panels is for precise data; the bottom row for interval-censored data with $w = 5$.

In Table 3.7 we give the Wald intervals of the small GSS data set for the precise and interval-censored data. Note that the Wald formulas for α_1 and α_2 based on the interval-censored data caused the left endpoint of the 95% intervals to extend below zero, so we reset these values to zero as α_1 and α_2 must be positive.

Table 3.7. Wald Intervals Based on the GSS Small Data Set and IC with $w_i = 5 \ \forall \ i$

		Precise	Interval Censored
Param	Truth	95% Wald	95% Wald
β_1	1.79	(1.41,2.33)	(1.09,2.21)
β_2	-0.41	$(-1.39, 0.19)$	$(-1.33, 0.42)$
λ_1	6.00	(3.50,9.45)	(2.36,8.07)
λ_2	4.00	(1.27,5.85)	(1.29,5.32)
α_1	1.50	(0.44,2.00)	(0.00,3.32)
α_2	2.50	(0.72,4.08)	(0.00,5.05)

Similarly to our investigation in Section 2.4.2, with this example we do not seek to demonstrate cases in which the Wald approximation will always fail, or will always provide an appropriate interval estimate. We have simply outlined a particular data scenario with relatively small sample size for which the Wald approximations on some of the parameters proved to be suspect. Based on these findings we do not recommend the Wald interval approximations as a general tool in the frequentist approach to the regression problem.

3.5 Prior Structures

Before we proceed with further analyses, we explore various prior structures for $\mathbf{r}$ and $\boldsymbol{\beta}$ in model (3.6) and consider their effects on posterior inference. Our Bayesian inference thus far has followed the convention used by Ntzoufras in his GSS data analysis, as introduced in Section 3.3.

3.5.1 Uniform Prior on Dispersion Parameter

First, as discussed in Section 2.5.1, placing a gamma$(0.001, 0.001)$ on the dispersion parameter r in (3.6) induces a heavy prior probability on $r = 0$. This prior can have an unintended effect on the posterior, especially for smaller sample sizes. One suggested alternative prior (Spiegelhalter *et al.* 2004, pp. 168-177) for a dispersion parameter is a uniform$(0, B)$ where the upper bound B is chosen so as to stabilize posterior features of interest, such as credible interval width or a specified posterior probability. We illustrate this approach using the small GSS data set depicted in Figure 3.7. Our trial values of B range from 1 to 20 in increments of 0.5. We plot the marginal posterior means for the components of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$ as a function of B. For this example the posterior means stabilize at $B = 4.5$; thus, our choice of prior in this case would be a uniform$(0, 4.5)$ placed on r_1 and r_2. We compare the use of the gamma(0.001,0.001) on r_1 and r_2 with this common uniform prior in Table 3.8: "Prior 1" (marked as a "1" in the table) indicates the gamma(0.001,0.001) on r_1 and r_2 with diffuse normals on the components of $\boldsymbol{\beta}$, while "Prior 2" (marked as a "2" in the table) indicates the uniform(0,4.5) on r_1 and r_2, again with diffuse normals on the components of $\boldsymbol{\beta}$.

Table 3.8. Posterior Results Based on Gamma and Uniform Priors on r_1 and r_2

Par	Truth	Mean 1	Mean 2	SD 1	SD 2
β_1	1.79	1.89	1.89	0.25	0.24
β_2	-0.41	-0.57	-0.58	0.44	0.41
λ_1	6.00	6.86	6.83	1.82	1.71
λ_2	4.00	4.04	3.94	1.81	1.49
α_1	1.50	1.37	1.22	0.47	0.42
α_2	2.50	2.75	2.42	1.04	0.90
Par	Truth	CS 1	CS 2	Error 1	Error 2
β_1	1.79	(1.43,2.42)	(1.45, 2.39)	0.0019	0.0018
β_2	-0.41	$(-1.40, 0.34)$	$(-1.37, 0.26)$	0.0033	0.0033
λ_1	6.00	(4.17,11.27)	(4.25, 10.88)	0.0142	0.0125
λ_2	4.00	(1.95,8.17)	(2.02,7.55)	0.0117	0.0113
α_1	1.50	(0.68,2.49)	(0.62,2.22)	0.0033	0.0029
α_2	2.50	(1.30,5.27)	(1.15,4.61)	0.0068	0.0068

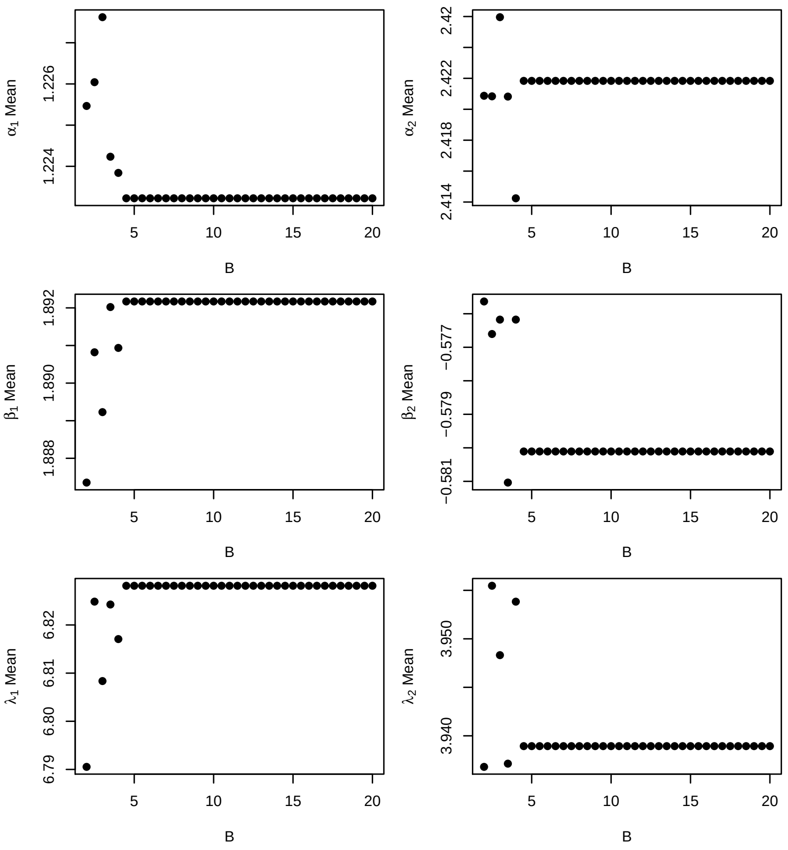

Figure 3.7: Posterior means for α_1 (top left), α_2 (top right), β_1 (center left), β_2 (center right), λ_1 (bottom left), and λ_2 (bottom right), calculated for increasing values of B for $r_j \sim \text{uniform}(0, B)$, $j = 1, 2$.

The final two columns in Table 3.8 list the MCMC error based on the two different prior structures; we see the difference in MCMC errors is mostly negligible. Convergence based on each prior structure seems to take roughly the same amount of time, as indicated by the Gelman-Rubin-Brooks plots in Figure 3.8. Prior 1 results are plotted in the six panels on the left, and Prior 2 results are plotted in the six panels on the right.

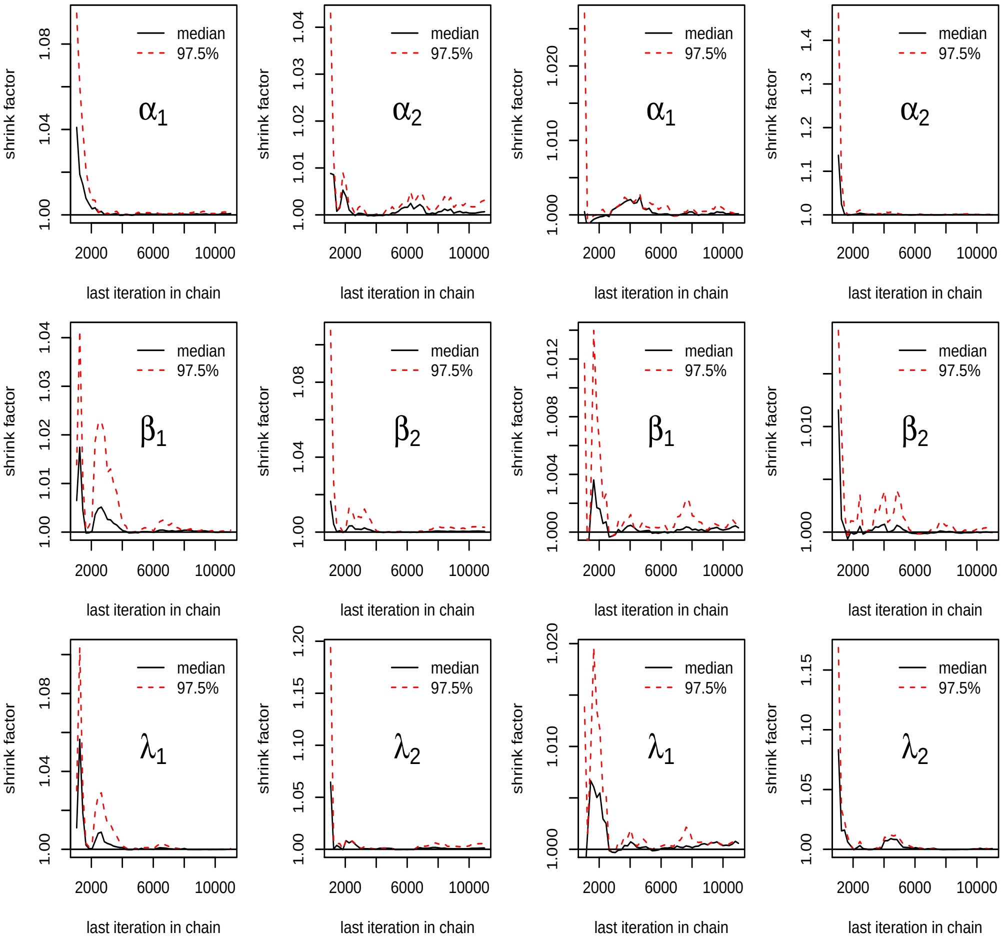

Figure 3.8: Gelman-Rubin statistic medians (solid black) and 97.5& percentiles (dashed red) for α_1 and α_2 (top row), β_1 and β_2 (center row), and λ_1 and λ_2 (bottom row). The left six panels represent Gelman-Rubin statistics based on gamma(0.001,0.001) priors for r_1 and r_2; the right six panels represent Gelman-Rubin statistics based on uniform(0,4.5) priors for r_1 and r_2.

The posterior means for each parameter remain roughly the same for these two prior structures, with the most notable difference in the α_2's: 2.75 for Prior 1 and 2.42 for Prior 2. Practically, if $\lambda_2 = 4$ this would mean a dispersion index (see Section 2.1) $DI = 12.00$ based on Prior 1 and $DI = 10.68$ based on Prior 2. The differences based on the prior structures can mostly be seen in the 95% credible sets, which are narrower across the board when we use the uniform prior structure (Prior 2) on r_1 and r_2.

3.5.2 Cauchy Prior on Regression Coefficients

We now consider the prior structure on the regression coefficients β_1 and β_2. The choice of a diffuse normal prior on regression coefficients is common when prior data or knowledge are not available; however, Gelman *et al.* (2008) suggest a Cauchy distribution alternative which they describe as "weakly informative". They suggest first scaling binary variables to have mean zero, and nonbinary variables to have mean zero and standard deviation 0.5. Then they place a Cauchy prior with center zero and scale 2.5 on each of the regression coefficients. In the Poisson case—or in our negative binomial case with log link—this indicates a prior belief that effects would not be greater than 5 on the logarithmic scale.

To illustrate use of this Cauchy prior, we implement model (3.6), using the small GSS data set described in Section 3.4. We set $\beta_j \sim \text{Cauchy}(0, 2.5)$ and $r_j \sim \text{Uniform}(0, 4.5)$ for $j = 1, 2$. We leave the gender variable unscaled, as there are an even number of men and women in the data set. We give the posterior results in Table 3.9 under the "Prior 3" columns. Prior 1 results are as listed in Table 3.8 and represent the diffuse normal priors on the components of $\boldsymbol{\beta}$ with Gamma(0.001,0.001) priors on r_1 and r_2.

In this example, the use of the Cauchy(0,2.5) prior on the components of $\boldsymbol{\beta}$ did not seem to improve precision beyond what had been achieved by using a uniform

Table 3.9. Posterior Results Based on Three Prior Structures

Par	Truth	Prior 1 Mean	Prior 3 Mean	Prior 1 SD	Prior 3 SD
β_1	1.79	1.89	1.86	0.25	0.23
β_2	-0.41	-0.57	-0.53	0.44	0.40
λ_1	6.00	6.86	6.63	1.82	1.62
λ_2	4.00	4.04	4.02	1.81	1.49
α_1	1.50	1.37	1.23	0.47	0.42
α_2	2.50	2.75	2.42	1.04	0.90
Par	Truth	Prior 1 CS	Prior 3 CS		
β_1	1.79	(1.43,2.42)	(1.42, 2.35)		
β_2	-0.41	$(-1.40, 0.34)$	$(-1.29, 0.28)$		
λ_1	6.00	(4.17,11.27)	(4.13, 10.46)		
λ_2	4.00	(1.95,8.17)	(2.06,7.63)		
α_1	1.50	(0.68,2.49)	(0.61,2.23)		
α_2	2.50	(1.30,5.27)	(1.14,4.56)		

on r_1 and r_2. The most noticeable difference came in the reduction in standard deviation for λ_1, with a slight reduction in width of the 95% credible set for λ_1.

3.5.3 Conditional Means Prior on Regression Coefficients

Now suppose we have prior knowledge in the form of expert opinion, prior data, or both, that we wish to incorporate into the prior structure on the components of $\boldsymbol{\beta}$. One method for incorporating such knowledge is through the use of conditional means priors (CMPs), as introduced by Bedrick *et al.* (1996). These priors are based on the notion that it is easier to formulate prior information in terms of mean values of the dependent variable conditioned on specified fixed values of the covariates, than to formulate prior information about regression coefficients. For example, suppose we are concerned with a population in which sexually transmitted disease is prevalent. Previous studies of this population provide estimates of the expected number of intercourses for men and women, λ_1 and λ_2, in this population. The CMP can make use of such information as well as expert opinion on these expected counts.

The CMP works in this manner: first, using expert opinion, previous data, or both, we construct a prior on $\tilde{\mathbf{m}} = (\tilde{m}_1, \ldots, \tilde{m}_p)'$, where each $\tilde{m}_i$ is the mean response at selected covariates $\tilde{\mathbf{x}}_i, i = 1, \ldots, p$. Let $\tilde{\mathbf{X}}$ be the $p \times p$ nonsingular matrix with $\tilde{\mathbf{x}}_i'$ as its ith row. For a link function, g, and a matrix, $\mathbf{M} = [m_{ij}]$, let $\mathbf{G} = [g(m_{ij})]$. For the inverse link, g^{-1}, define $\mathbf{G}^{-1}$ similarly. Then we have

$$\tilde{\mathbf{m}} = \mathbf{G}^{-1}(\tilde{\mathbf{X}}'\boldsymbol{\beta}),$$

and

$$\boldsymbol{\beta} = \tilde{\mathbf{X}}^{-1}\mathbf{G}(\tilde{\mathbf{m}}).$$

If we place a proper prior, π_0, on $\tilde{\mathbf{m}}$, then the induced prior on $\boldsymbol{\beta}$ must also be proper and has the form

$$\pi(\boldsymbol{\beta}) = \pi_0(\mathbf{G}^{-1}(\tilde{\mathbf{X}}\boldsymbol{\beta}))|d\mathbf{G}^{-1}(\tilde{\mathbf{X}}\boldsymbol{\beta})|,$$

where $d\mathbf{G}^{-1}(\tilde{\mathbf{X}}\boldsymbol{\beta})$ is the matrix of partial derivatives. If we can elicit the $\tilde{m}_i$'s independently, the CMP is given by

$$\pi_0(\tilde{\mathbf{m}}) = \prod_{i=1}^{p} \pi_{0i}(\tilde{m}_i),$$

and the resulting induced prior is given by

$$\begin{aligned}\pi(\boldsymbol{\beta}) &= \frac{\prod_{i=1}^{p} \pi_{0i}(g^{-1}(\tilde{\mathbf{x}}_i'\boldsymbol{\beta}))}{|\tilde{\mathbf{X}}^{-1}| \prod_{i=1}^{p} \dot{g}(\tilde{m}_i)} \\ &\propto \frac{\prod_{i=1}^{p} \pi_{0i}(g^{-1}(\tilde{\mathbf{x}}_i'\boldsymbol{\beta}))}{\dot{g}(g^{-1}(\tilde{\mathbf{x}}_i{}'\boldsymbol{\beta}))}.\end{aligned}$$

Here $\dot{g}(a)$ denotes $dg(a)/da$.

CMPs are related to data augmentation priors (DAPs). The prior on the conditional mean values has the functional form of the likelihood in a DAP, and it is elicited by specifying "prior observations" along with weights for those observations. Bedrick *et al.* (1996) note that, in the Poisson regression model with log link, if a gamma prior is placed on the mean parameter then this corresponds to the induced prior on $\boldsymbol{\beta}$ being a DAP. The corresponding CMP has distribution gamma$(\tilde{w}_i\tilde{y}_i, \tilde{w}_i)$. Here $\tilde{y}_i$ is a prior estimate of the Poisson mean for specified values of the covariates, denoted $\tilde{\mathbf{x}}_i$, and $\tilde{w}_i$ is viewed as the prior number of observations associated with $\tilde{y}_i$. The weight $\tilde{w}_i$ may be a fractional number. Because we use the log link in our negative binomial models, in their role representing overdispersed Poisson data, we will use this choice of CMP.

For example, suppose prior opinion or data suggest an average of 6 and 3 intercourses for men and women, respectively, in a population of interest. If an expert is willing to assign a weight of 5 observations for both men and women, then we have

$$\lambda_1 \sim \text{gamma}(6 \times 5, 5), \quad \lambda_2 \sim \text{gamma}(3 \times 5, 5).$$

Here we have $\tilde{\mathbf{x}}_1 = (1\ 0)'$, $\tilde{\mathbf{x}}_2 = (1\ 1)'$, $\tilde{m}_1 = \tilde{y}_1 = 6$, and $\tilde{m}_2 = \tilde{y}_2 = 3$ so that the induced prior on $\boldsymbol{\beta}$ is the bivariate distribution given by

$$\begin{aligned}
\pi(\boldsymbol{\beta}) =& \frac{\prod_{i=1}^{2} \pi_{0i}(\exp(\tilde{\mathbf{x}}_i'\boldsymbol{\beta}))}{|\tilde{\mathbf{X}}^{-1}| \prod_{i=1}^{2} 1/(\tilde{m}_i)} \\
=& \frac{\frac{5^{30}}{\Gamma(30)}[\exp(\tilde{\mathbf{x}}_1'\boldsymbol{\beta})]^{30-1} \exp\{-5\exp(\tilde{\mathbf{x}}_1'\boldsymbol{\beta})\} \frac{5^{15}}{\Gamma(15)}[\exp(\tilde{\mathbf{x}}_2'\boldsymbol{\beta})]^{15-1} \exp\{-5\exp(\tilde{\mathbf{x}}_2'\boldsymbol{\beta})\}}{\left|\left(\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right)^{-1}\right| (1/6)(1/3)} \\
=& 18\frac{5^{30}}{\Gamma(30)}[\exp(\beta_1)]^{30-1} \exp\{-5\exp(\beta_1)\} \frac{5^{15}}{\Gamma(15)}[\exp(\beta_1+\beta_2)]^{15-1} \\
& \times \exp\{-5\exp(\beta_1+\beta_2)\}.
\end{aligned}$$

In the next section, we will see further examples of this joint prior.

3.5.4 Comparison of the Three Prior Structures

To illustrate the effects of the $N(0, 1000)$, Cauchy$(0, 2.5)$ and conditional means priors for $\boldsymbol{\beta}$, we use the GSS small data set created in Section 3.4 and plot the perspective plots and contours of the priors, likelihood, and posteriors for $\boldsymbol{\beta}$ in Figure 3.9. Note that $\boldsymbol{\alpha}$ is fixed in these cases at the "true" generating values $\alpha_1 = 1.5$, $\alpha_2 = 2.5$.

The diffuse normal and the Cauchy priors on β_1 and β_2 yield a posterior with a shape similar to that of the likelihood, while the CMP method—in which we assume to have obtained informative expert opinion on λ_1 and λ_2—drastically changes the posterior contour from the original shape of the likelihood. This is the desired effect as we sought to "inform" our analysis, based on prior information, with the conditional means prior.

Of course, in our analysis we might primarily be interested in the expected counts for men and women, λ_1 and λ_2, as well as the dispersion parameters α_1 and α_2. Figure

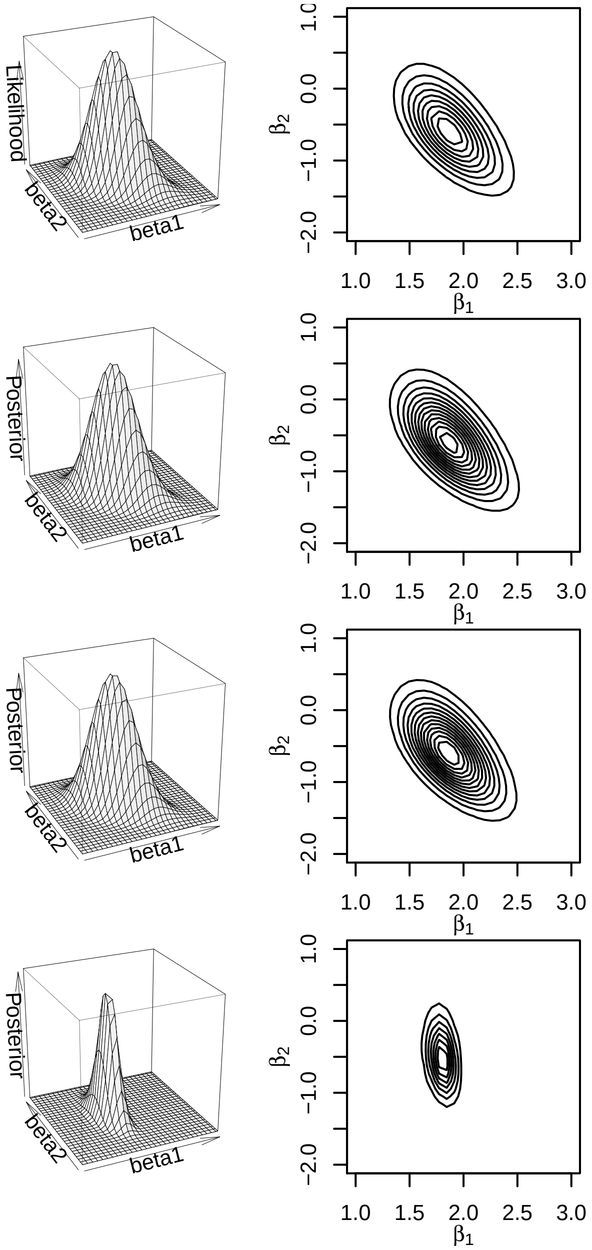

Figure 3.9: Likelihood for (β_1, β_2) based on the precise small GSS data set and corresponding posteriors for (β_1, β_2) based on three different priors on the components of $\boldsymbol{\beta}$: N(0,1000) (second row), Cauchy(0,2.5) (third row), and conditional means priors from $\lambda_1 \sim \text{gamma}(30, 5)$, $\lambda_2 \sim \text{gamma}(15, 5)$ (bottom row). Surface plots are given in the left column; contour plots of the surfaces are given in the right column.

3.10 shows the effect of these three choices in $\boldsymbol{\beta}$ priors on the posteriors for λ_1, λ_2, α_1, and α_2. In each case, we maintained uniform(0,4.5) priors on r_1 and r_2. We also examine the effect when using interval-censored data: we interval-censor each of the data points with $w_i = 5 \forall\ i$, and the position of the true value y_i randomly placed within $[j_i, k_i]$.

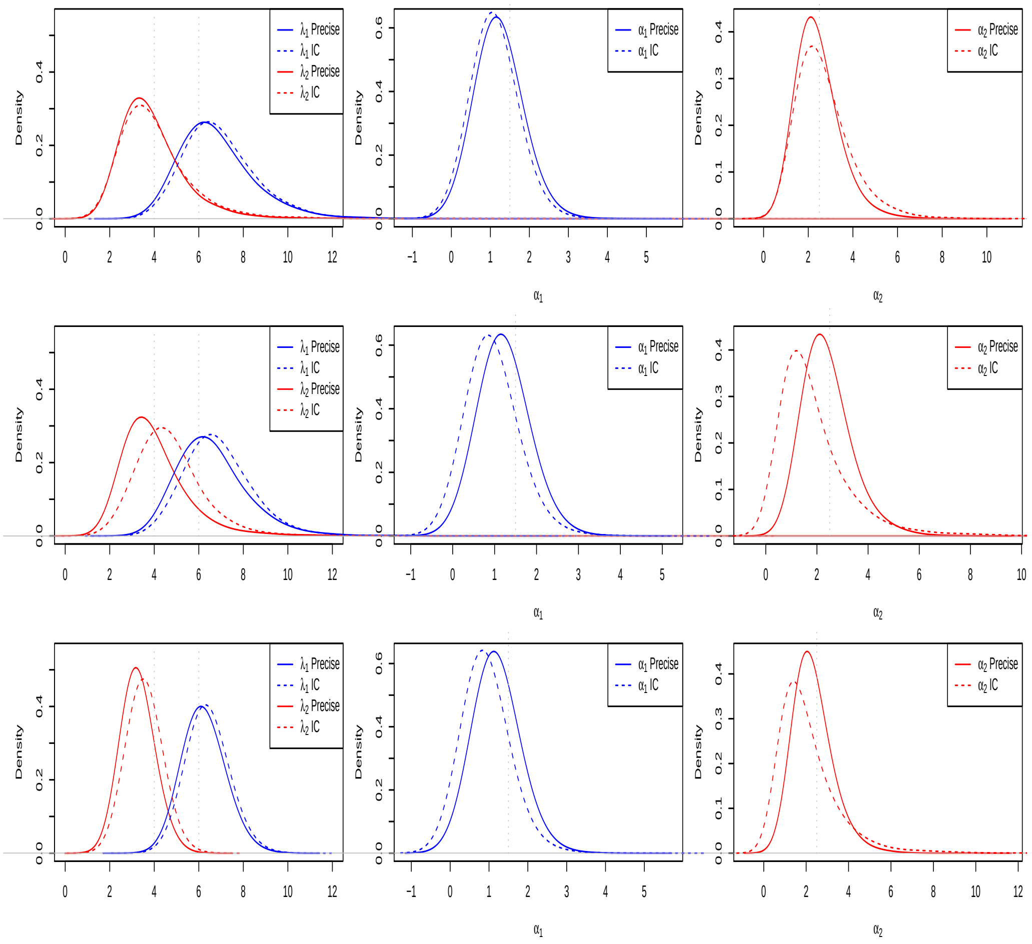

Figure 3.10: Posterior densities for λ_1 and λ_2 (left column), α_1 (center column), and α_2 (right column) based on three different priors on β_1 and β_2: N(0,1000) (top row), Cauchy(0,2.5) (center row), and conditional means priors from $\lambda_1 \sim \text{Gamma}(30, 5)$, $\lambda_2 \sim \text{Gamma}(15, 5)$ (bottom row). Precise results are given by solid lines; interval-censored results are given by dashed lines.

The posteriors on α_1 and α_2 remain roughly the same regardless of marginal prior choice on the components of $\boldsymbol{\beta}$, in the precise-count case. However, the posteriors on α_1 and α_2 resulting from the interval-censored data seem sensitive to the choice of priors for the components of $\boldsymbol{\beta}$. Less diffuse choices of $\boldsymbol{\beta}$ prior (the Cauchy and the CMP) seem to pull the posterior distributions on α_1 and α_2 further to the left, indicating smaller values of α_1 and α_2, meaning less data dispersion. *A priori* the parameters α_1 and α_2 are most directly affected by the prior on r_1 and r_2, as $\alpha_1 = 1/r_1$ and $\alpha_2 = 1/r_2$. With a uniform(0,4.5) prior on r_1, the induced prior on α_1 is also relatively uniform in this range (and likewise for r_2 and α_2). Then the effects on the posterior distributions in Figure 3.10 may be due to the behavior of λ_1 and λ_2 in the interval-censored case: because the dispersion parameter is inversely related to the expected count, we expect smaller α_1 and α_2 for larger λ_1 and λ_2.

The λ_1 and λ_2 marginal posteriors are very similar for the precise and interval-censored cases, when using $\mathrm{N}(0, 1000)$ priors on the components of $\boldsymbol{\beta}$. These posteriors differ a bit more for the Cauchy and CMP priors. When we use the CMPs, the resulting posteriors on the components of $\boldsymbol{\lambda}$ are less diffuse and indicate greater differentiation between λ_1 and λ_2. The posterior for λ_1 shrinks from the true value of 4 and toward our elicited “prior belief” value of 3. Figure 3.10 displays the posteriors for λ_1 and λ_2. In Table 3.10 we consider the posterior probability, using the different priors, that $\beta_2 < 0$, implying a smaller expected count for females.

Table 3.10. Posterior Inference for β_2 Based on Several Prior Choices

$\boldsymbol{\beta}$ Priors	Case	$P(\beta_2 < 0 \vert \mathbf{d}_n)$	95% CS for β_2
Normal	Precise	0.9194	$(-1.37, 0.25)$
Normal	IC	0.8846	$(-1.21, 0.32)$
Cauchy	Precise	0.9101	$(-1.29, 0.29)$
Cauchy	IC	0.8699	$(-1.17, 0.34)$
CMP	Precise	0.9984	$(-1.13, -0.20)$
CMP	IC	0.9948	$(-1.07, -0.16)$

The CMP, normal, and Cauchy priors yield 95% posterior intervals for both the precise and interval-censored cases that do not contain zero, implying a difference between men and women in reported number of sexual intercourses. The corresponding values of $P(\beta_2 < 0|\mathbf{d}_n)$ are also very high.

3.6 Likelihood Behavior

In this section, we examine the behavior of the likelihood for the interval-censored negative binomial model. Our treatment here is similar to that in Section 2.3 in the non-regression context. We focus on the likelihood contours for $\boldsymbol{\beta}$ and corresponding contours for $\boldsymbol{\lambda}$, as these are of primary interest. For these examples we will fix $\alpha_1 = 1.5$ and $\alpha_2 = 2.5$ based on the GSS data, as described in Section 3.1.

Given the results in Section 2.3, it is not surprising that here too the interval-censored widths will impact the shape of the likelihood. We use our small GSS data set from Section 3.4, and examine the effects on the components of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ when increasing the interval-censored widths. In Figure 3.11 we begin with $w_i = 1 \; \forall \; i$, increase to $w_i = 2 \forall i$, and lastly examine the effects when $w_i = 5 \; \forall \; i$. The precise likelihood contours are in solid blue, and the interval-censored likelihood contours are in dashed red. The profile likelihoods for β_1 and β_2, and λ_1 and λ_2, are plotted along the top and right axes. As expected, wider intervals in the interval-censored data cause the likelihood contours to differentiate further from those based on the precise counts.

As anticipated, when we increase the percentage of interval-censored observations, the likelihood departs markedly from the precise case. This is illustrated in Figure 3.12, in which we take the $n = 50$ precise data and interval-censor 25% of the points, then 50% of the points, and finally 100% of the points. For all censored data, we have $w_i = 2$.

As in the non-regression case, increasing sample size diminishes the differences between precise and interval-censored likelihoods. In Figure 3.13 we increase n from 30 to 50 to 100 (with equal numbers of male and female responders in each set).

The location of the true y_i within $[j_i, k_i]$ will also have an impact on the shape of the likelihood, as we saw in Chapter 2. In Figure 3.14 we plot the extremes of $y_i = j_i \; \forall \; i$

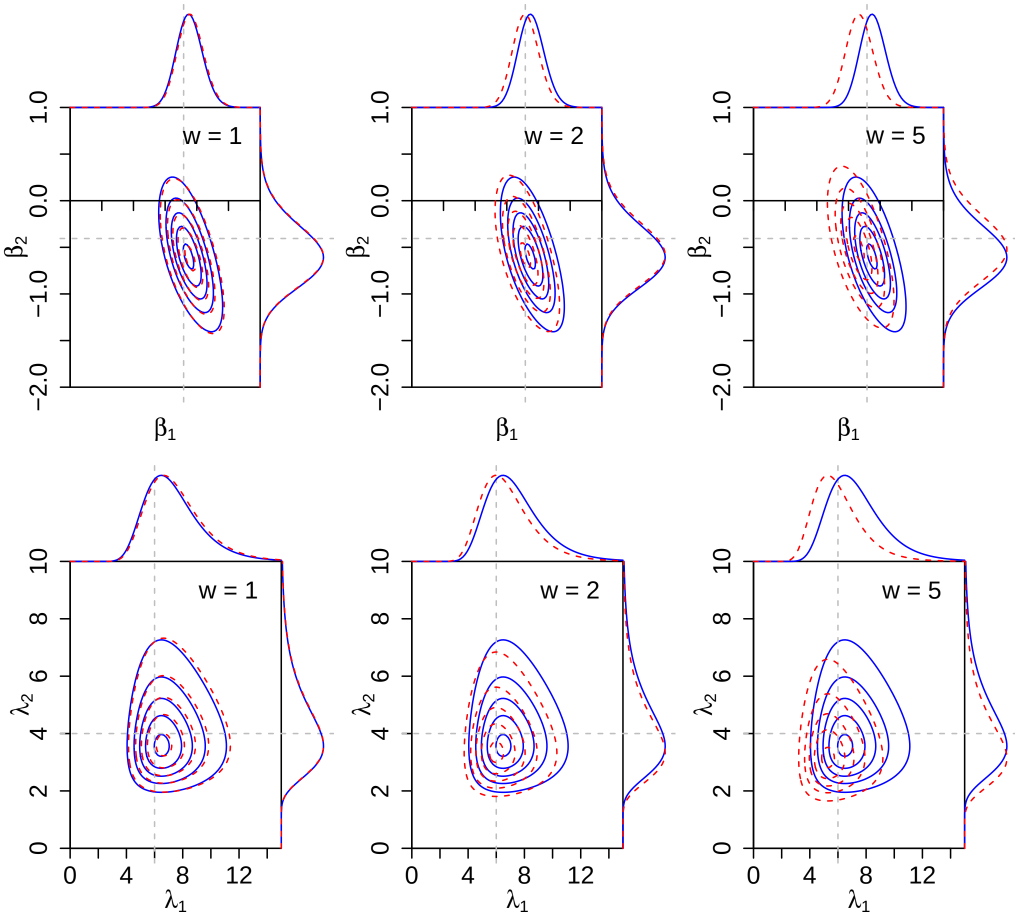

Figure 3.11: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, based on the small GSS data set. Contours for β_1 and β_2 given on top row; contours for λ_1 and λ_2 given on bottom row. Profile likelihoods of β_1 (or λ_1) on upper external axes and profile likelihoods of β_2 (or λ_2) on outer right axes. The left column has $w = 1$; center column $w = 2$; and right column $w = 5$.

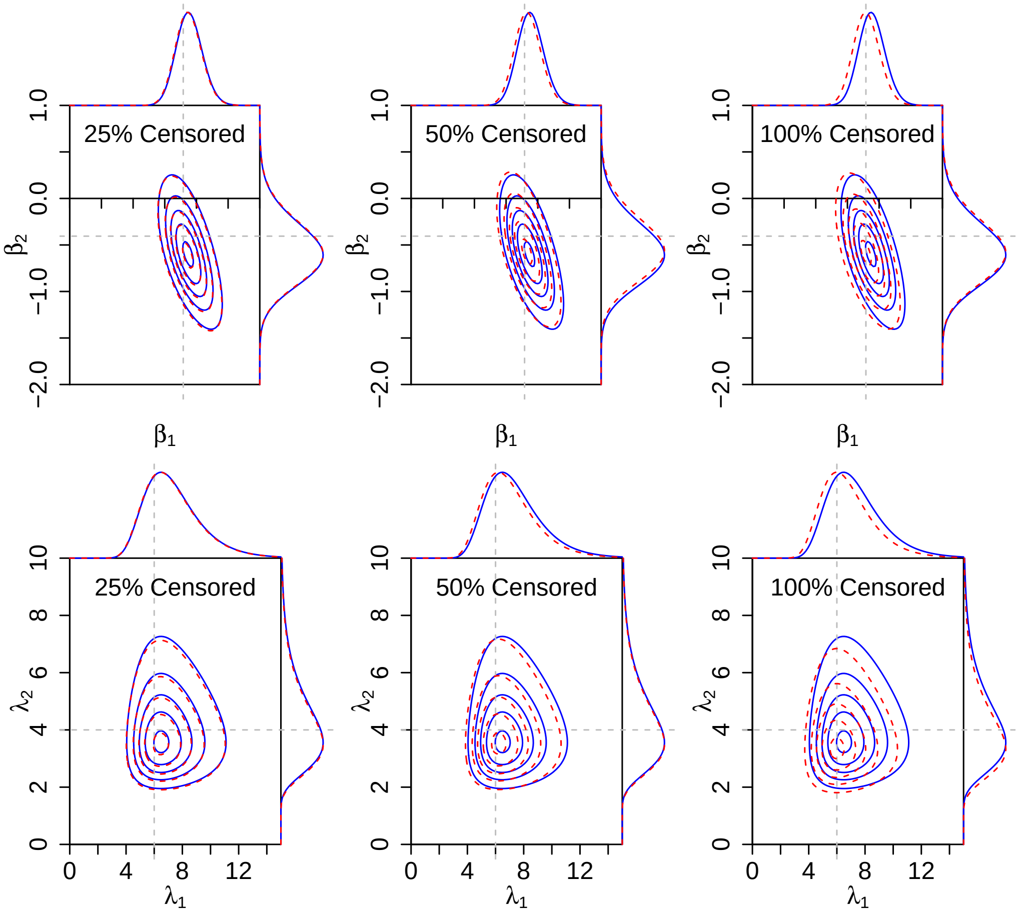

Figure 3.12: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, based on the small GSS data set. Interval-censored widths $w = 2$ for all interval-censored data. Contours for β_1 and β_2 given on top row; contours for λ_1 and λ_2 given on bottom row. Profile likelihoods of β_1 (or λ_1) on upper external axes and profile likelihoods of β_2 (or λ_2) on outer right axes. The left column has 25% censoring; center column 50% censoring; and right column 100% censoring.

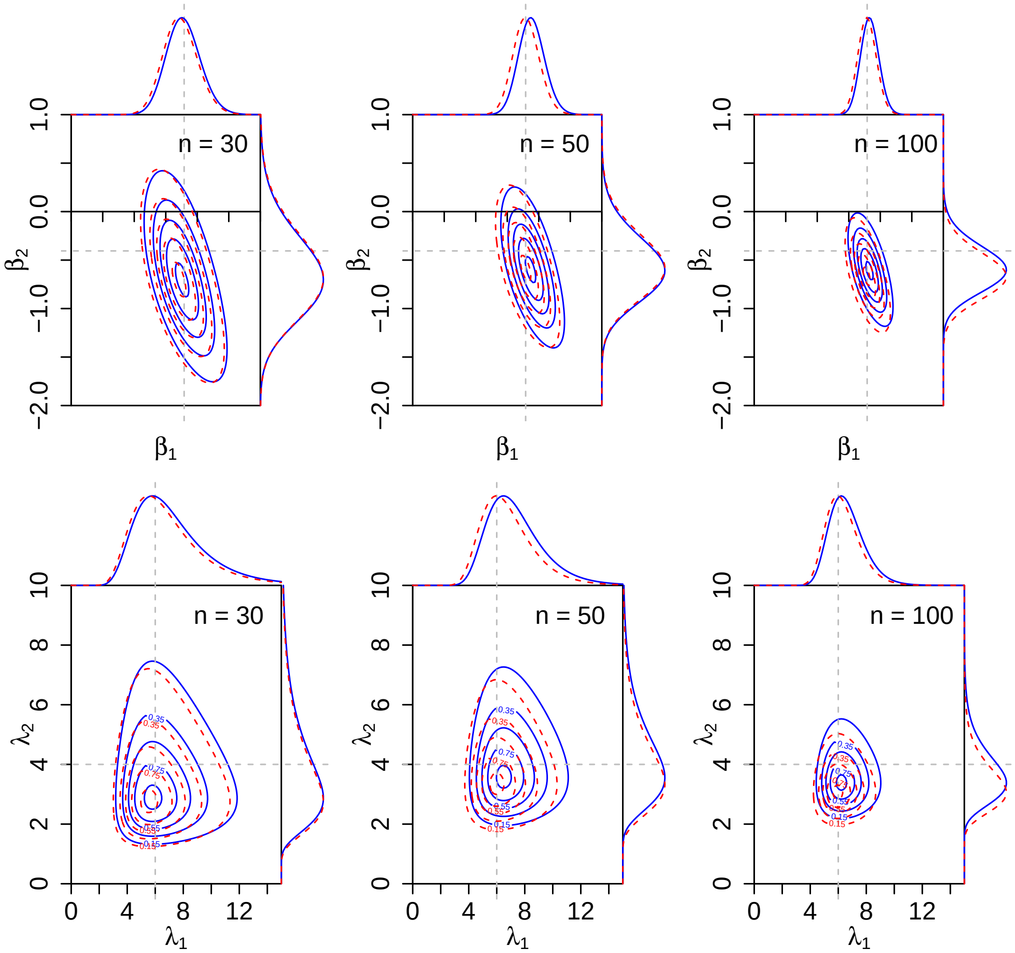

Figure 3.13: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data; data generated based on true underlying parameter values behind the small GSS data set. Interval-censored widths $w = 2$ for all interval-censored data. Contours for β_1 and β_2 given on top row; contours for λ_1 and λ_2 given on bottom row. Profile likelihoods of β_1 (or λ_1) on upper external axes and profile likelihoods of β_2 (or λ_2) on outer right axes. The left column has $n = 30$; center column $n = 50$; and right column $n = 100$.

and $y_i = k_i \; \forall \; i$, which would correspond to all respondents giving high estimates or low estimates with their interval reports, respectively. In the center plot we illustrate the effect of having $y_i = (j_i + k_i)/2 \; \forall \; i$; that is, the precise value is at the center of each interval. For each plot, $w_i = 6 \; \forall \; i$. However, in the second and third plot groups, setting $w_i = 6$ would cause several of the j_i to fall below zero, so those j_i values are reset to zero and this causes only 74% of the data to be interval-censored. In the center plot group, k_i is set to $y_i \times 2$ whenever j_i is set to zero, so as to keep y_i centered in $[j_i, k_i]$.

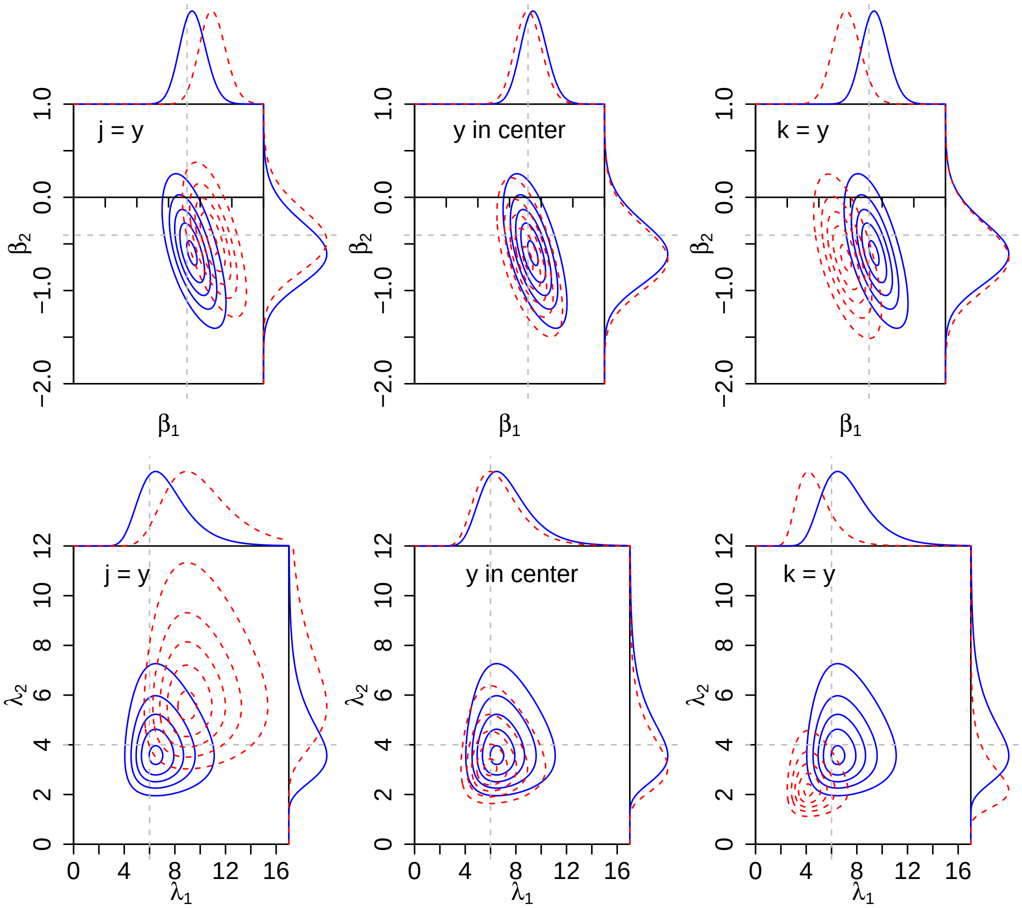

Figure 3.14: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data, based on the small GSS data set. Interval-censored widths $w = 6$ for all interval-censored data. Contours for β_1 and β_2 given on top row; contours for λ_1 and λ_2 given on bottom row. Profile likelihoods of β_1 (or λ_1) on upper external axes and profile likelihoods of β_2 (or λ_2) on outer right axes. The left column has $y_i = j_i \; \forall \; i$; center column $j_i = y_i - 3$ and $k_i = y_i + 3$; and right column $y_i = k_i \; \forall \; i$.

In the extremes where $y_i = j_i$ or $y_i = k_i \; \forall \; i$, there is clearly an effect on the concentration and location of the likelihood. If y_i happens to be centered within each interval $\forall \; i$, the interval-censored likelihoods are very similar to the precise likelihoods—even though the interval width is 6. When the precise value falls at the lower interval endpoints, the profile likelihoods for λ_1 and λ_2 are concentrated at higher values and exhibit corresponding higher variability in the likelihood. There is an opposite effect when $y_i = k_i \; \forall \; i$. The changes in precision might be explained by the relationship between $\boldsymbol{\lambda}$ and the variance of the responses, which we examined in Chapter 2.

As mentioned, for these examples we have fixed $\alpha_1 = 1.5$ and $\alpha_2 = 2.5$, based on rough estimates from the original GSS data, in order to view the $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ likelihoods as two-dimensional. Now in Figure 3.15 we vary the underlying fixed values of $\boldsymbol{\alpha}$ in order to show the effect of the dispersion parameter in this regression context.

The likelihoods maintain their shapes as $\boldsymbol{\alpha}$ increases. The value of $\boldsymbol{\lambda}$ in the regression context is most directly dependent on $\boldsymbol{\beta}$, so this is not surprising. However, there is a decrease in the concentration of the likelihood for increasing $\boldsymbol{\alpha}$. This is the expected relationship, as we saw in Section 2.1 that the variance of the counts is $\lambda + \alpha\lambda^2$, so higher α indicates higher variability and thus more dispersion in the likelihood. The differentiation between the interval-censored case and the precise case seems to remain constant for increasing values of α_1 and α_2.

In this section we have explored the effects on the likelihood contours of various factors that should be considered in the interval-censored problem, for several individual data sets. In the following section we provide a more in-depth exploration of several of these factors and the estimation of the corresponding regression models, with a simulation experiment.

3.7 Simulation Studies

In this section we present a simulation experiment to compare the interval-censored negative binomial regression model (3.2) to several alternative methods of handling interval-censored data. Within a simulation replication we generate precise count responses, $\mathbf{y}$,

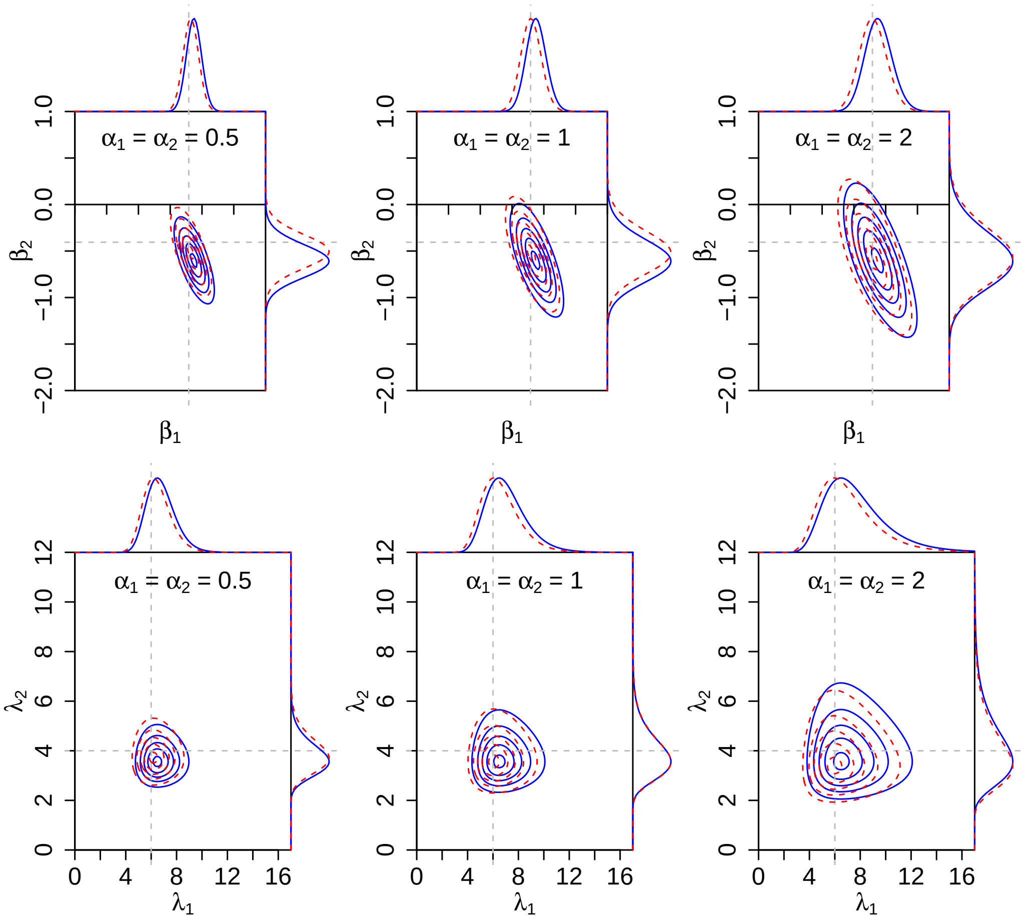

Figure 3.15: Likelihood contours for interval-censored (red dashed curves) and precise (blue solid curves) data; data generated based on true parameter values underlying the small GSS data set. Interval-censored widths $w = 2$ for all interval-censored data. Contours for β_1 and β_2 given on top row; contours for λ_1 and λ_2 given on bottom row. Profile likelihoods of β_1 (or λ_1) on upper external axes and profile likelihoods of β_2 (or λ_2) on outer right axes. The left column has α_1 and α_2 fixed at 0.5; center column α_1 and α_2 fixed at 1; and right column α_1 and α_2 fixed at 2.

based on a single binary covariate, x_i, and underlying true values of the regression coefficients, β_1 and β_2, and dispersion parameters, α_1 and α_2. We analyze the traditional negative binomial regression model for this data, then interval-censor each y_i to produce $\mathbf{d}_n = \{[j_i, k_i]\}_{i=1}^{n}$. Similar to our study in Section 2.7, we compare four different methods of handling the interval-censored counts $\mathbf{d}_n$:

(1) Apply the interval-censored regression model (3.2) to the data

(2) Fit the precise regression model (3.1) using the "mean method": use the midpoint of each $[j_i, k_i]$, rounded to the nearest whole number, as the precise count

(3) Fit the precise regression model (3.1) taking $y_i \equiv j_i \ \forall \ i$

(4) Fit the precise regression model (3.1) taking $y_i \equiv k_i \ \forall \ i$

We refer to these as the interval-censored, mean, lower, and upper methods, respectively, throughout our study. We choose parameter values similar to those estimated from the GSS example introduced in Section 3.1. As in Section 2.7, we choose the dispersion parameters so as to induce a reasonable amount of variability on $\mathbf{y}$. Highly variable data produces parameter estimates that are widely dispersed and that exhibit little differentiation among the methods. The design points are summarized in Table 3.11. Case 1 and Case 2 take λ_1 and λ_2 far apart at 7 and 3. Case 3 and Case 4 bring λ_1 and λ_2 closer together at 7 and 6. The values of α_1 and α_2 remain constant across the cases. We use a sample size of $n = 30$ for all cases, with 15 observations of each covariate value within each data set.

In the simulation cases we also consider the positioning of the precise count y_i within the interval $[j_i, k_i]$. This has an impact on the relative shape and location of the joint likelihoods as demonstrated in Figure 3.14. Just as in Section 2.7, we control the tendency of y_i to be near j_i or k_i for a given design point as noted in the last column of Table 3.11. After generating each y_i we randomly select its position, p_i, within the interval Since $w_i = 3 \ \forall \ i$ in each design point, p_i can take on a value of 1, 2, 3, or 4. In those design points for which y_i tends toward j_i, the selection probabilities are weighted more heavily toward lower values of p_i. Similarly, the probabilites are weighted more heavily toward higher values of p_i for design points in which y_i tends toward k_i.

Table 3.11. Design Points for Regression Simulation Study

Data Feature	Case 1	Case 2	Case 3	Case 4
w	3.00	3.00	3.00	3.00
λ_1	7.00	7.00	7.00	7.00
α_1	0.50	0.50	0.50	0.50
Var 1	31.50	31.50	31.50	31.50
SD 1	5.61	5.61	5.61	5.61
λ_2	3.00	3.00	6.00	6.00
α_2	1.50	1.50	1.50	1.50
Var 2	16.50	16.50	60.00	60.00
SD 2	4.06	4.06	7.75	7.75
β_1	1.95	1.95	1.95	1.95
β_2	-0.56	-0.56	-0.15	-0.15
y_i tendency	j_i	k_i	j_i	k_i

For each design point we run a *WinBUGS* analysis on the precise, interval-censored, mean method, lower method, and upper method data. The priors for each Bayesian model are $\pi \sim \text{beta}(1,1)$ and $r \sim \text{uniform}(0,500)$. We bring this uniform upper bound down from 1000 as employed in Section 2.7 due to the larger upper bound yielding convergence issues in the regression context. For each of the four design points we run 50 iterations. From each iteration we record the posterior medians to serve as point estimates, as well as the posterior standard deviations and equal-tailed 95% credible sets.

Just as in Section 2.7, we focus our comparisons for the five methods based on point and interval estimates for the parameters. The plotted simulation results display summaries of bias, credible set width, and credible set coverage. The average bias is represented by the distances of the point estimate means from the true values of the parameters indicated by the horizontal lines in each plot. We also plot the simulation means of the lower and upper 95% interval endpoints along with shaded boxes. The shaded boxes represent one standard deviation above and below average posterior means and interval limits. Coverage is demonstrated by the inclusion of the true parameter value in a design point's plotted summary interval. Additionally in this regression context, we examine the significance of the coefficient on the covariate, β_2. Significance is a frequentist concept, but we adapt it here to test the hypothesis of $\beta_2 \neq 0$ in the Bayesian paradigm.

We use the term *significant* in this section to refer to a regression coefficient that is more likely to be nonzero given the data model. To measure this in our simulation studies, we record the proportion of times that zero is included in credible sets for β_2. We also plot a reference line for zero in the summary plots on β_2 as a means of visualizing significance. Results for Cases 1 and 2 from Table 3.11 are plotted in Figures 3.16–3.18.

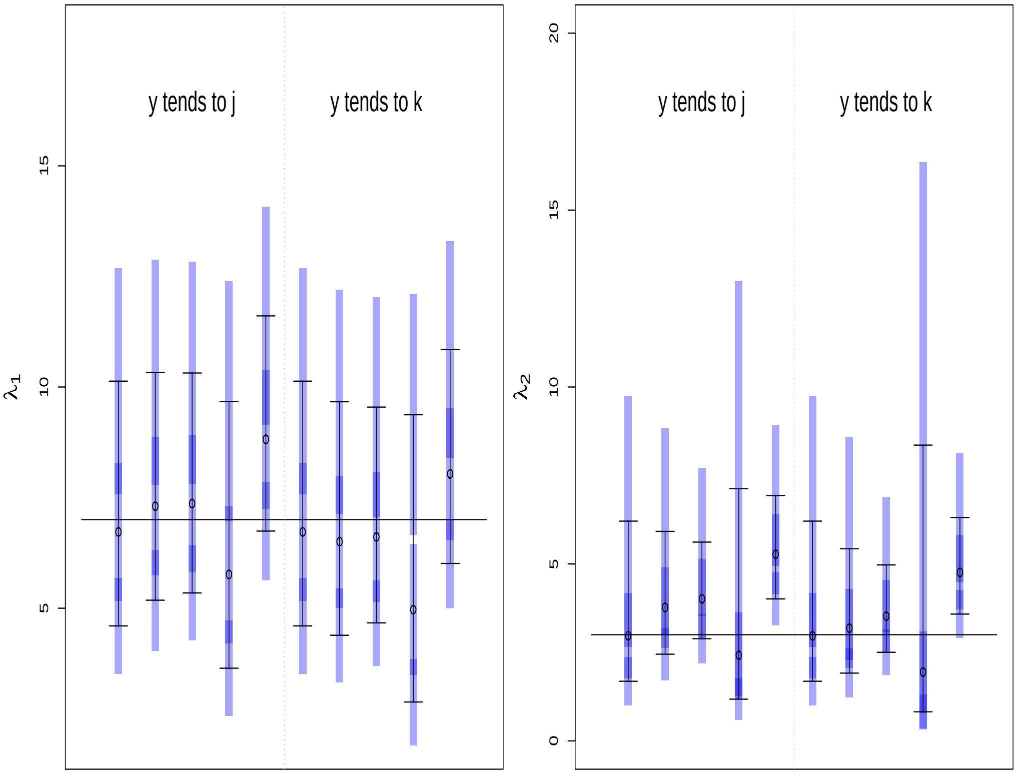

Figure 3.16: Bayesian simulation results for λ_1 (left panel) and λ_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$ and $\lambda_2 = 3$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

In Figure 3.16, the interval-censored and mean methods perform similarly in estimating λ_1, with small bias and good coverage. The lower and upper methods are more biased than the other methods, regardless of the tendency of y_i. There is more differentiation among the methods in estimating λ_2. When y_i tends toward k_i, the interval-censored method out-performs the other three methods, as it has the smallest bias and its average credible set is more centered over the true parameter value than those credible sets based on the other methods for interval-censored data. When y_i tends toward k_i, the lower method might seem comaprable to the interval-censored method as its average bias is similar and its average credible set has better coverage, but this comes at a cost of a much wider average credible set width.

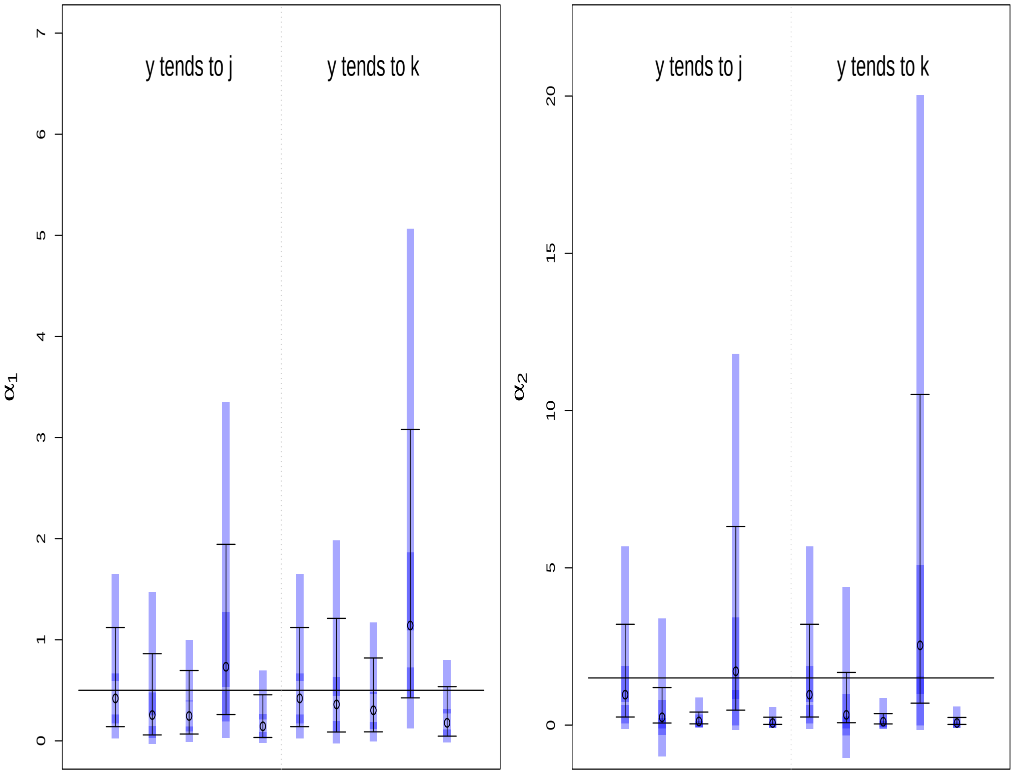

Figure 3.17: Bayesian simulation results for α_1 (left panel) and α_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$, $\lambda_2 = 3$, $\alpha_1 = 0.5$, and $\alpha_2 = 1.5$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

In Figure 3.17, the interval-censored method seems superior in estimating α_1 as it has less bias and better coverage than the other methods. As for α_2, the interval-censored average credible set upper bound falls below the true parameter value when y_i tends toward j_i, indicating poor coverage, but the mean and upper methods provide worse coverage for this parameter. Furthermore, though the lower method intervals do contain the true value and this method shows less bias than the interval-censored method, this method again gives very wide interval estimates.

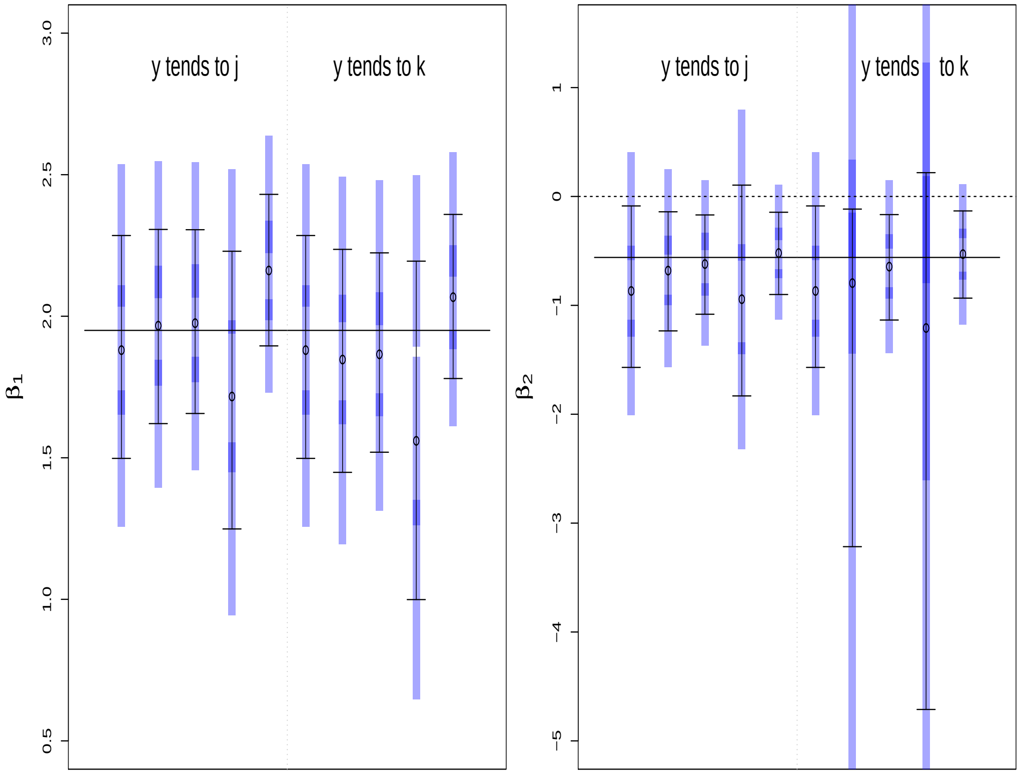

Figure 3.18: Bayesian simulation results for β_1 (left panel) and β_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$, $\lambda_2 = 3$, $\beta_1 = 1.95$ and $\beta_2 = -0.56$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

The interval-censored and mean methods perform very similarly in estimating β_1 (Figure 3.18), with small bias and interval estimates centered around the truth. The lower and upper methods yield more bias. The interval-censored, mean, and upper methods perform similarly in estimating β_2 when y_i tends toward j_i, while the lower method gives a wider average credible set, greater average bias, and it is the only method in this case whose average credible set contains zero (incorrectly indicating that β_2 is not significant). When y_i tends to k_i, there is large variability among the simulation replications for the interval-censored and lower methods. The mean and upper methods perform well, with small bias and shorter intervals that are centered around the truth. Again for this case, the lower method is the only method that would indicate that β_2 is not significant—incorrectly implying that the covariate does not have a significant effect on the expected count. The widths and coverages of the interval estimates for all six parameters are summarized in Tables 3.12–3.15. We also report significance values for β_2 in Table 3.15.

Table 3.12. Coverage for $\lambda_1(= 7)$ and $\lambda_2(= 3)$

Par	Tend	Method	Est	Avg	SD	Cov	Tend	Est	Avg	SD	Cov
λ_1	j_i	Prec	6.73	5.53	1.79	0.88	k_i	6.73	5.53	1.79	0.88
λ_1	j_i	IC	7.31	5.15	1.81	0.92	k_i	6.51	5.28	1.85	0.86
λ_1	j_i	Mean	7.37	4.97	1.66	0.92	k_i	6.61	4.88	1.71	0.86
λ_1	j_i	Lower	5.76	6.03	2.11	0.82	k_i	4.97	6.49	2.16	0.78
λ_1	j_i	Upper	8.82	4.86	1.62	0.54	k_i	8.04	4.83	1.65	0.80
λ_2	j_i	Prec	2.97	4.53	3.22	0.84	k_i	2.97	4.53	3.22	0.84
λ_2	j_i	IC	3.77	3.47	2.62	0.74	k_i	3.19	3.51	2.87	0.86
λ_2	j_i	Mean	4.01	2.73	1.50	0.60	k_i	3.52	2.47	1.37	0.74
λ_2	j_i	Lower	2.42	5.95	5.65	0.82	k_i	1.95	7.54	7.85	0.74
λ_2	j_i	Upper	5.27	2.92	1.38	0.04	k_i	4.76	2.73	1.26	0.26

For Case 3 and Case 4 in Table 3.11, simulation summary plots look very similar to those presented for Case 1 and Case 2. The essential difference in these cases is the increase in the true value of λ_2 from 3 to 6, so that λ_1 and λ_2 are close. Refer to Appendix D for the full simulation summaries on Cases 3 and 4. In this section we will present results for β_2, as estimates for this parameter differed the most between the first two cases

Table 3.13. Coverage for $\alpha_1 (= 0.5)$ and $\alpha_2 (= 1.5)$

Par	Tend	Method	Est	Avg	SD	Cov	Tend	Est	Avg	SD	Cov
α_1	j_i	Prec	0.42	0.98	0.43	0.88	k_i	0.42	0.98	0.43	0.88
α_1	j_i	IC	0.26	0.80	0.54	0.72	k_i	0.36	1.12	0.70	0.84
α_1	j_i	Mean	0.25	0.63	0.24	0.80	k_i	0.30	0.73	0.28	0.84
α_1	j_i	Lower	0.73	1.68	1.19	0.86	k_i	1.14	2.66	1.71	0.66
α_1	j_i	Upper	0.14	0.42	0.20	0.46	k_i	0.18	0.49	0.21	0.60
α_2	j_i	Prec	0.97	2.95	2.19	0.70	k_i	0.97	2.95	2.19	0.70
α_2	j_i	IC	0.25	1.13	2.05	0.22	k_i	0.33	1.60	2.56	0.32
α_2	j_i	Mean	0.12	0.37	0.39	0.04	k_i	0.11	0.33	0.41	0.02
α_2	j_i	Lower	1.71	5.84	4.97	0.82	k_i	2.54	9.82	8.83	0.70
α_2	j_i	Upper	0.07	0.23	0.28	0.00	k_i	0.07	0.22	0.29	0.00

Table 3.14. Coverage for $\beta_1 (= 1.95)$

Tend	Method	Est	Avg	SD	Cov	Tend	Est	Avg	SD	Cov
j_i	Prec	1.88	0.79	0.17	0.88	k_i	1.88	0.79	0.17	0.88
j_i	IC	1.97	0.69	0.17	0.92	k_i	1.85	0.79	0.19	0.86
j_i	Mean	1.98	0.65	0.12	0.92	k_i	1.87	0.70	0.13	0.84
j_i	Lower	1.72	0.98	0.26	0.82	k_i	1.56	1.20	0.29	0.78
j_i	Upper	2.16	0.54	0.11	0.58	k_i	2.07	0.58	0.11	0.78

Table 3.15. Coverage for $\beta_2 (= -0.56)$

Tend	Method	Est	Avg	SD	Cov	Sig
j_i	Prec	-0.87	1.48	0.38	0.86	0.56
j_i	IC	-0.68	1.09	0.30	0.88	0.64
j_i	Mean	-0.62	0.91	0.14	0.90	0.68
j_i	Lower	-0.94	1.94	0.56	0.82	0.44
j_i	Upper	-0.52	0.76	0.12	0.92	0.70
k_i	Prec	-0.87	1.48	0.38	0.86	0.56
k_i	IC	-0.80	3.10	12.89	0.84	0.60
k_i	Mean	-0.64	0.97	0.14	0.88	0.66
k_i	Lower	-1.21	4.93	17.81	0.82	0.38
k_i	Upper	-0.53	0.80	0.11	0.92	0.68

and the second two cases. The summary plots for β_2 based on $\lambda_1 = 7$ and $\lambda_2 = 6$ are given in Figure 3.19.

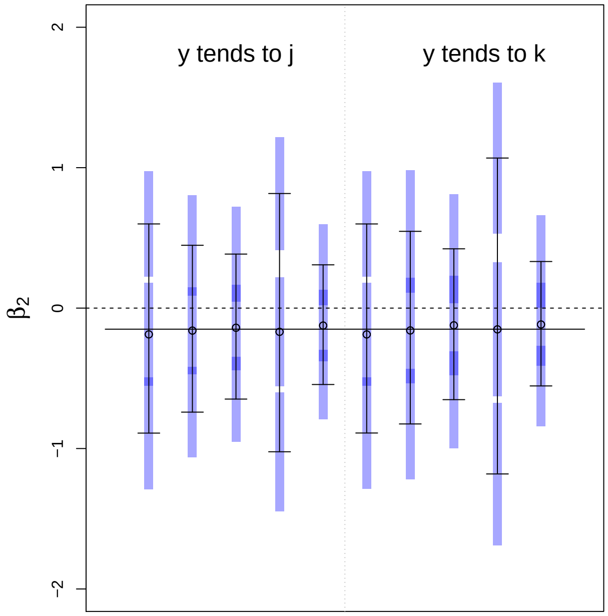

Figure 3.19: Bayesian simulation results for β_2, based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$ and $\lambda_2 = 6$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

When we compare the simulation summaries in Figure 3.19 to 3.18, we see a reduction in bias for several of the methods, as all average posterior medians are very close to the true β_2 for the $\lambda_2 = 6$ case. There is some variation in the interval widths, with the lower method producing the widest intervals and the upper method producing the narrowest intervals. The average widths based on the interval-censored method come closest to matching the average widths in the precise case. As expected, when we bring λ_1 and λ_2 closer together, we see that the significance of β_2 is greatly reduced in the case, with the average credible sets tending toward being centered around zero. The performances of the intervals for β_2 are summarized in Table 3.16.

Table 3.16. Coverage for $\beta_2(= -0.15)$

	y_i tends to j_i					y_i tends to k_i				
Method	Est	Avg	SD	Cov	Sig	Est	Avg	SD	Cov	Sig
Prec	-0.19	1.49	0.26	0.80	0.04	-0.19	1.49	0.26	0.80	0.04
IC	-0.16	1.19	0.23	0.74	0.08	-0.16	1.37	0.31	0.74	0.12
Mean	-0.14	1.03	0.16	0.60	0.12	-0.12	1.07	0.15	0.60	0.18
Lower	-0.17	1.84	0.31	0.84	0.02	-0.15	2.25	0.47	0.82	0.06
Upper	-0.12	0.85	0.14	0.50	0.14	-0.12	0.89	0.14	0.52	0.18

3.7.1 Discussion

In this simulation study, we compared a precise-count negative binomial regression model to the interval-censored model developed in this chapter, as well as three alternative methods of handling the interval-censored data. We quantified our comparisons by the average bias of the point estimates across simulation replications, the average width and coverage of the interval estimates for the parameters, and the "significance" of the regression coefficient for the single binary covariate.

Some of our findings were similar to those in Section 2.7, in which we detailed the simulation study for the non-regression context. We found that the interval-censored method performed similarly to the mean method for estimates on β_1 and λ_1, just as there were similarities between these methods for some design points in Section 2.7. However, the interval-censored method was superior to the mean method, in terms of bias and interval coverage of the true parameter, for λ_2, α_1, and α_2. The lower method appeared to have reduced bias and better coverage in estimating α_1 and α_2, but at a cost of much wider credible sets.

When we increased λ_2 to be close to λ_1, we saw more agreement among the four methods in estimating β_2, with results for the other parameters remaining similar to those in the first two cases. Overall, for the cases under our consideration, the interval-censored method performed at least as well as (and sometimes better than) the other methods for handling interval-censored data.

In future work, we will expand on the design points explored here. One important factor to consider in future simulation designs is the use of additional covariates. Our examples in this chapter focused on the case of a single binary covariate, and our simulation study was an extension of those examples. However, it will be important to observe the effects on parameter estimation when additional binary covariates are added into the regression model, as well as continuous covariates. We will also explore various sample sizes beyond the $n = 30$ selected for this study, as well as a broader range of underlying true parameter values.

CHAPTER FOUR

Application: Health-Related Quality of Life Data

We now apply our interval-censored negative binomial model to actual data from a recent study. Kobau *et al.* (2007) performed a study on a population of California adult survey respondents in which they analyzed the effect of having epilepsy on outcomes such as self-rated health, quality of life, and access to health care. The data for their study were obtained from the California Health Interview Survey (CHIS), which is a random-dial telephone interview conducted every two years since 2001, with around 50,000 California households per year. The survey is conducted by the UCLA Center for Health Policy Research in partnership with many public and private organizations. The survey questions included vary somewhat from year to year, and four survey questions regarding epilepsy were included in the 2003 and 2005 surveys but excluded from the 2001, 2007 and 2009 surveys. Kobau *et al.* (2007) used the 2003 CHIS data for their study. The four questions regarding epilepsy from this survey are as follows:

> QA03_43 Has a doctor ever told you that you have seizure disorder or epilepsy? (if 'No', 'Refused', or 'Don't Know', skip the remaining three questions)
>
> QA03_44 Are you now taking any medicine to control your seizure disorder or epilepsy?
>
> QA03_45 How many seizures of any type have you had in the last three months?
>
> QA03_46 During the past month, to what extent has your epilepsy or its treatment interfered with normal activities like working, school, or getting together with family or friends? Would you say
>
> - Not at all
> - Slightly
> - Moderately

- ∗ Quite a bit or
- ∗ Extremely?

Some of the survey variables are not made available for public use based on their potential for leading to public identification of a respondent, or their confidential nature (i.e., variables regarding sexual behavior). CHIS administrators masked the responses to QA03_45, regarding the number of seizures in the past month, from the public use file. Kobau *et al.* used this variable in their study, but we will not use it here given its unavailability for public use.

Our dependent variable of interest will be the total number of unhealthy days experienced by the respondents in the past 30 days. The values actually reported by respondents in the CHIS survey were number of physically unhealthy days, and separately the number of mentally unhealthy days in the last 30 days. The total number of days that the respondent felt unwell (either physically or mentally) was not reported directly and has the potential to be interval-censored. For example, a person who reports 4 physically unhealthy days and 2 mentally unhealthy days may have been unwell for 4 total days, 6 total days, or a partial overlap of 5 total days. Then that respondent would have an interval-censored value for their total number of unhealthy days given by $[4, 6]$. There is a logical maximum of 30 unhealthy days based on the nature of the questions. For example, if a person reported 20 physically unhealthy days and 20 mentally unhealthy days, the maximum total number of unhealthy days is capped at 30 and their interval would be $[20, 30]$. Often there are precise counts as well in the data set, particularly when the respondent reports a positive number of unhealthy days in only one category. For example, if the respondent experienced 5 physically unwell days and 0 mentally unwell days, their summary score would be a precise count of 5 total unwell days.

The CDC first created these survey questions as a measure of health-related quality of life (HRQOL), and this organization recommends the upper-bound method (adding physically and mentally unhealthy days as a summary index of overall healthy days, with a logical maximum of 30). Kobau *et al.* (2007) analyzed the relationship between the epilepsy questions above and physically or mentally unhealthy days separately, without an analysis of the summary index. The authors found that “Adults who have had epilepsy reported about twice as many mentally unhealthy days, physically unhealthy days, and activity limitation days as those with no history of epilepsy, and these differences were all significant.” Their study controlled for age, gender, race/ethnicity, income, and chronic disease comorbidity. The authors define chronic disease comorbidity as an indicator for whether the respondent was ever told they had cancer or heart disease, and/or had current diabetes or current asthma. There are 42,044 total responses in the data set; our analysis will focus on a small subset of this data. Table 4.1 gives the frequencies and percentages of the full data demographic information, categorized by whether the respondent has ever been told by a doctor that they have epilepsy or seizure disorder.

4.1 Data Subset

Suppose a researcher is interested in subjects who are epileptic, but otherwise quite healthy. Then we can obtain a subset from the full data by identifying those who have been told by a doctor that they have epilepsy or seizure disorder but do not have any of the listed comorbidities. The result is a subset of the data with $n = 309$ observations. A subset of this data is obtained by using another variable from the survey, the respondent’s self-rated health. The question is worded “In general would you say you health is:” with response options of “Excellent,” “Very good,” “Good,” “Fair,” or “Poor.” We will consider only those subjects who rated their health as excellent or very good – in general, they consider themselves to be healthy and do not have a major comorbidity of interest, but their seizure disorder may contribute to an increase in number of unhealthy days. The result is our final working data set with $n = 129$.

For our analysis we will compare the interval-censored model with the two extremes of precise models. If P_i and M_i represent the reported number of physically unwell and mentally unwell days, respectively, then the minimum total number of unhealthy days for a given respondent will be

$$j_i \equiv \max(P_i, M_i). \tag{4.1}$$

The maximum total number of unhealthy days will be

Table 4.1. Descriptive Variables in HRQOL-Epilepsy Study

	Ever told had epilepsy			
	Yes		No	
Variable	N	%	N	%
Total	550	1.2	41,494	98.8
Age (years)				
18-34	94	25.8	9,611	33.3
35-44	129	22.8	8,511	21.7
45-64	241	39.6	14,790	30.3
65+	86	11.8	8,582	14.8
Gender				
Male	225	46.9	17,252	49.0
Female	325	53.1	24,242	51.0
Race/Ethnicity				
White, non-Hispanic	352	55.9	24,877	48.7
Black, non-Hispanic	40	7.0	2,510	6.1
Hispanic	96	26.5	8,674	30.8
Asian/Pacific Isl.	25	5.2	3,893	12.0
American Indian	7	1.6	342	0.7
Other Race	30	3.7	1,198	1.7
Annual Household Income				
$< \$25,000$	239	45.3	11,276	29.3
$\$25,000 - \$49,999$	118	22.1	10,155	23.7
$\geq \$50,000$	193	32.6	20,063	47.1
Comorbidities	241	43.8	11,717	28.2
Cancer	85	15.5	4,642	11.2
Heart Disease	85	15.5	3,578	8.6
Diabetes	55	10.0	2,849	6.9
Asthma	99	18.0	3,488	8.4

$$k_i \equiv \min(P_i + M_i, 30), \tag{4.2}$$

where 30 is a logical maximum due to the nature of the questions posed. For example, if a respondent reported 20 physically unwell days and 15 mentally unwell days, the minimum total unwell days would indicate complete overlap of the physical and mental unwell days: $\max(20, 15) = 20$. The maximum total unwell days would be $\min(20 + 15, 30) = 30$. This latter total illustrates the method suggested by the CDC. Our interval-censored model incorporates this uncertainty, recording an interval from minimum j_i to maximum k_i, $[20, 30]$. In this way, we have three datasets based on the initial data, with each dataset treating the interval-censored data in a different manner. The first two data sets employ the precise-case negative binomial likelihood, and the third data set will make use of our interval-censored negative binomial likelihood (3.2).

We will regress the unhealthy days response on a single binary covariate, similar to our work in Chapter 3. Our covariate of interest will be an age indicator, $x_i = 0$ if the respondent is aged $18 - 44$ and $x_i = 1$ if the respondent is 45 or older. In our data set there are 68 respondents (53%) in the 45+ age group. This cutoff represents the midpoint of the four age groups used in the Kobau *et al.* (2007) study.

We plot the frequencies of total number of unhealthy days, calculated by (4.1) as well as (4.2), separated by age group in Figure 4.1. We refer to the calculation method in (4.1) as the "lower" method, and in (4.2) as the "CDC" method.

Note that, especially for the $18 - 44$ age group, there is differentiation in the calculated total number of unhealthy days between the lower and CDC methods. The interval-censored model will accomodate these discrepancies and retain the full amount of information we have for a given respondent. In the complete data set there are 21 subjects (16% of the respondents) for whom the lower count does not equal the CDC count, and thus those subjects have an interval-censored count. The means and variances for the groups are given in Table 4.2.

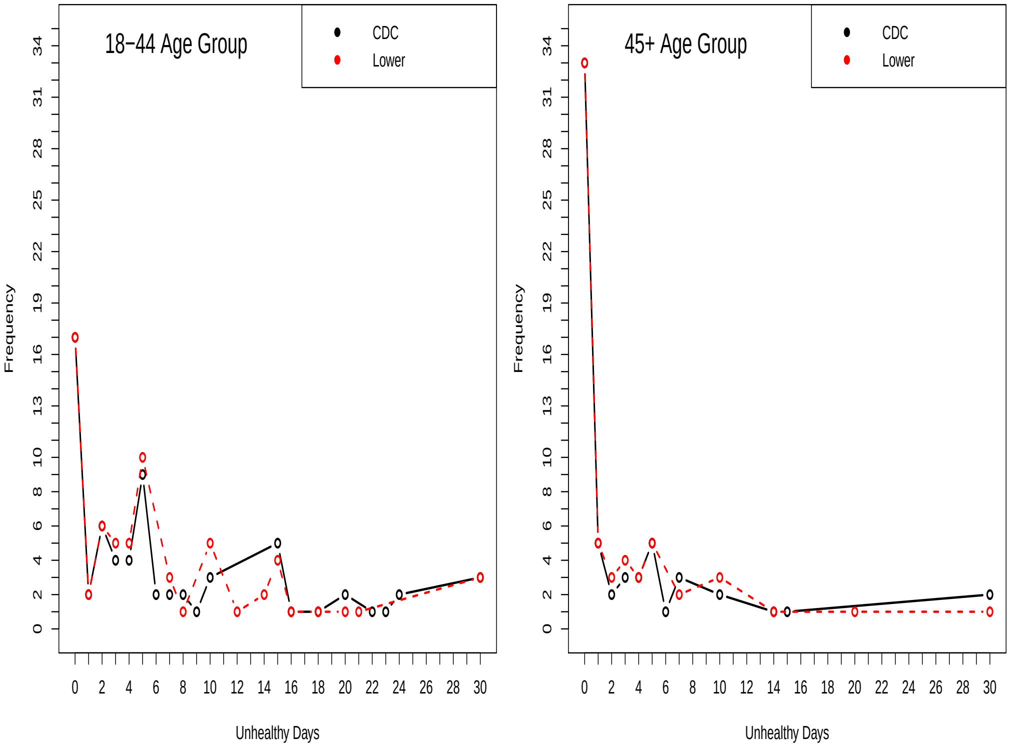

Figure 4.1: Frequencies of total number of unhealthy days, lower limit method (red) and CDC method (black). Frequences for the 18 − 44 age group are given in the left panel; frequencies for the 45+ age group are given in the right panel.

Table 4.2. Means and Variances for Age Groups, CDC vs. Lower Methods

	Lower		CDC	
Age	Mean	Variance	Mean	Variance
18 − 44	6.51	56.46	7.35	70.50
45+	2.75	27.62	3.13	36.98

Clearly there is overdispersion in the data. In the following section we will compare the Poisson and negative binomial fits to ensure that the negative binomial model is appropriate.

4.2 Frequentist Analysis

We first fit negative binomial and Poisson models to the precise-count data sets. The GLM for this data is given by

$$Y_i \sim \text{NB}(\lambda_i, \alpha_i),$$

where Y_i is the total number of unhealthy days, which falls in the interval $[j_i, k_i]$ with probability one. If $j_i = k_i$ then y_i is a precise observed count. The log-linear model is given by

$$\log(\lambda_i) = \beta_1 + \beta_2 x_i, \tag{4.3}$$

where λ_i is the expected total number of unhealthy days and x_i is an indicator for whether the subject is in age group 45+. We will let λ_a and λ_b denote the expected counts for patients in age groups $18 - 44$ and 45+, respectively, with $\lambda_a = \exp(\beta_1)$ and $\lambda_b = \exp(\beta_1 + \beta_2)$. Then α_a and α_b will represent the dispersion parameters for the two groups of subjects.

The Poisson model is given by $Y_i \sim \text{Poisson}(\lambda_i^p)$, with the log link as specified in (4.3). We fit only the precise-case Poisson model to the precise data sets. This is accomplished using the `glm` package in *R* (2012) and the negative binomial model with `glm.nb` (2012). The estimates from the `glm.nb` function are identical to those obtained from `ml.nb2` (altered to employ the Newton-Raphson method, code available upon request). Using `glm.nb` here allows for easier creation of the fitted model from the parameter estimates. In Figure 4.2 we plot the relative frequencies of the lower limit data on the left and the CDC summary data on the right, then the Poisson and negative binomial GLM fits to each precise data set. By inspection of these plots, it is evident that the Poisson model does not capture the shape of the observed data as well as the negative binomial. We will not consider the Poisson model further. Refer to Watson (2011) for a thorough treatment of the interval-censored Poisson regression model.

Table 4.3 summarizes the negative binomial fits for the two precise cases given in Figure 4.2, as well as the interval-censored negative binomial regression estimates. Here we have used our versions of the *R* functions `ml.nb2` and `ml.nb2.cens` for producing all estimates. Recall that 16% of the data are interval-censored. Note that the 95% confidence intervals for β_2 under all methods are below zero. This would imply that, in this data subset, those aged 45+ have a significantly lower number of reported total unhealthy days.

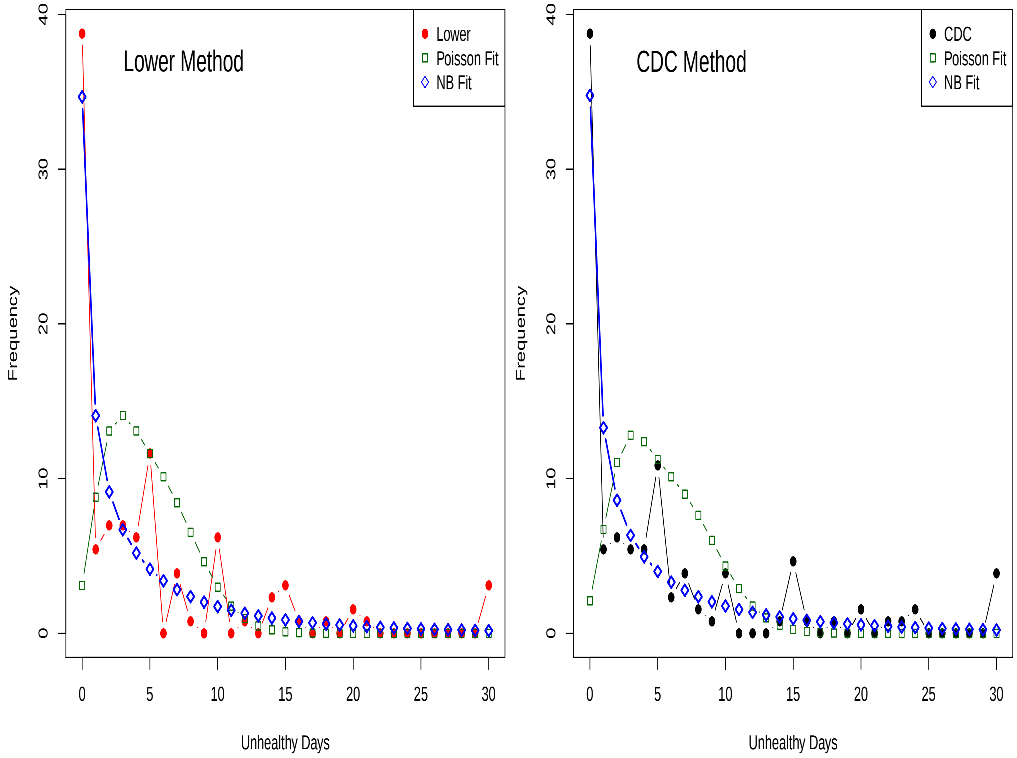

Figure 4.2: Negative binomial (blue diamond) and Poisson (green square) regression fits to the precise-case data methods of calculating total unhealthy days in the CHIS data subset: Lower method (left panel, red dots) and CDC method (right panel, black dots).

4.3 Bayesian Analysis

In the Bayesian approach to this problem, we choose a prior structure for the β's as well as the dispersion parameters. We employ the regression coefficient prior structure $\beta_j \sim \text{Cauchy}(0, 2.5)$ for $j = 1, 2$. As discussed in Section 3.5.2, this Cauchy prior is an alternative to the more typical diffuse normal distributions. For the dispersion parameters r_1 and r_2 we will employ a uniform$(0, B)$ prior as discussed in Section 3.5.1, suggested by Spiegelhalter *et al.* (2004, pp. 168-177). The value of B is driven by the particular data scenario and the posterior features of interest, so we must investigate this for our epilepsy data just as we investigated B for the GSS data in Section 3.5.1. This investigation is provided in Appendix E, on the CDC method precise data. Results were similar for the lower method and interval-censored data. We find that values of $B \geq 6$ produce only small changes in the posterior means and standard deviations for our parameters of interest, so

Table 4.3. Frequentist Estimates

Method	Param	Est	S.E.	95% Wald CI
Lower Precise	β_1	1.87	0.16	(1.57,2.18)
Lower Precise	β_2	-0.86	0.30	$(-1.46, -0.27)$
Lower Precise	λ_a	6.52	1.01	(4.53,8.50)
Lower Precise	λ_b	2.75	0.72	(1.34,4.16)
Lower Precise	α_a	1.49	0.31	(0.88,2.09)
Lower Precise	α_b	3.82	0.99	(1.88,5.76)
CDC Precise	β_1	1.99	0.16	(1.68,2.31)
CDC Precise	β_2	-0.85	0.31	$(-1.47, -0.24)$
CDC Precise	λ_a	7.35	1.18	(5.04,9.66)
CDC Precise	λ_b	3.13	0.84	(1.48,4.78)
CDC Precise	α_a	1.61	0.32	(0.97,2.25)
CDC Precise	α_b	4.10	1.04	(2.06,6.13)
Int. Censored	β_1	1.93	0.16	(1.62,2.24)
Int. Censored	β_2	-0.86	0.31	$(-1.47, -0.26)$
Int. Censored	λ_a	6.87	1.09	(4.74,9.00)
Int. Censored	λ_b	2.91	0.77	(1.39,4.42)
Int. Censored	α_a	1.54	0.32	(0.92,2.16)
Int. Censored	α_b	3.93	1.01	(1.95,5.92)

we set $B = 6$ in our Bayesian analyses. The model is summarized in Figure 4.3, with priors ultimately placed on r_1, r_2, β_1 and β_2.

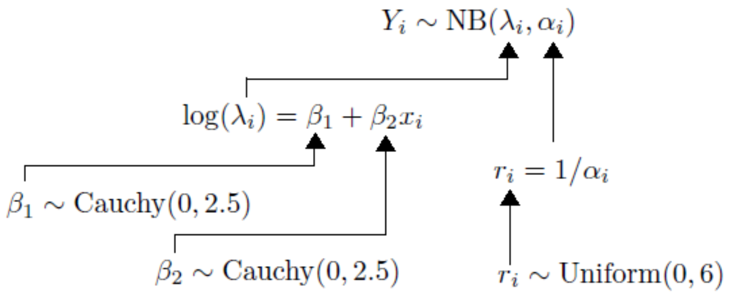
$Y_i \sim \text{NB}(\lambda_i, \alpha_i)$
$\log(\lambda_i) = \beta_1 + \beta_2 x_i$
$r_i = 1/\alpha_i$
$\beta_1 \sim \text{Cauchy}(0, 2.5)$
$\beta_2 \sim \text{Cauchy}(0, 2.5)$
$r_i \sim \text{Uniform}(0, 6)$

Figure 4.3. Schematic of the Bayesian model priors for the CHIS data subset.

We run the model in *WinBUGS* using a burn-in period of 1,000 iterations, with 10,000 updates after a thinning of 10. In Table 4.4 we present the posterior summaries for the three data sets.

The 95% equal-tailed credible intervals under all methods are to the left of zero, again indicating fewer unhealthy days reported for the 45+ group than for the $18-44$ group. The estimates for the components of $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}$ given by the posterior means all increase as we move from the lower method to the interval-censored method to the CDC method. Thus the interval-censored method seems to provide intermediate estimates between the two extremes of precise counts, as we would expect, without discarding information from the survey respondents.

In Figure 4.4 we plot the posterior densities of the expected counts for the two different age groups, λ_a and λ_b. These posteriors reflect the trend in Table 4.4, with those for λ_a and λ_b both shifting to the right as the data methods change from lower, to interval-censored, to CDC. Note also the increase in posterior dispersion, which can be explained by the increasing values of the components of both $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ in each set, and the direct relationship between the variance of the observed counts with these parameters, $V(Y_i) = \lambda_i + \alpha_i\lambda_i^2$. The plotted densities appear to indicate a difference between λ_a and λ_b under all three methods, but we can also quantify this appearance with the posterior probability $P(\beta_2 < 0)$ under each method. We calculate this by the proportion of values in the first chain for β_2 that fall below zero. These approximations are given in Table 4.5.

All three data sets indicate a very high posterior probability that $\beta_2 < 0$, in favor of those in the 45+ age group reporting a smaller number of unhealthy days than those in the $18-44$ age group. Other posterior inferences are easy to obtain. For example, we could estimate the difference, $\Delta_\lambda \equiv \lambda_a - \lambda_b$, addressing the decrease in unhealthy days from the younger to the older group. Posterior summaries for Δ_λ are given in Table 4.6.

We also analyze the posterior probability that λ_a exceeds λ_b by at least 3 days, or at least 4 days. These values are approximated by the proportion of values in the δ_λ chain that exceed the noted quantities, for each data method, in Table 4.7. Here we see more differentiation among data methods than in the probabilities in Table 4.5. The difference in probabilities is approximately 0.08 from the lower to CDC methods for $P(\lambda_a - \lambda_b > 3)$, and approximately 0.14 for $P(\lambda_a - \lambda_b > 4)$.

Table 4.4. Bayesian Estimates

Method	Param	Est	S.D.	95% CS
Lower Precise	β_1	1.87	0.15	(1.58,2.18)
Lower Precise	β_2	-0.81	0.31	$(-1.40, -0.19)$
Lower Precise	λ_a	6.56	1.02	(4.84,8.84)
Lower Precise	λ_b	2.99	0.84	(1.75,4.98)
Lower Precise	α_a	1.49	0.31	(0.97,2.21)
Lower Precise	α_b	3.85	1.03	(2.26,6.23)
CDC Precise	β_1	1.99	0.16	(1.68,2.31)
CDC Precise	β_2	-0.80	0.32	$(-1.40, -0.15)$
CDC Precise	λ_a	7.41	1.20	(5.39,10.10)
CDC Precise	λ_b	3.42	1.00	(1.98,5.83)
CDC Precise	α_a	1.61	0.33	(1.07,2.36)
CDC Precise	α_b	4.14	1.07	(2.46,6.60)
Int. Censored	β_1	1.92	0.16	(1.62,2.24)
Int. Censored	β_2	-0.81	0.31	$(-1.40, -0.17)$
Int. Censored	λ_a	6.92	1.11	(5.05,9.39)
Int. Censored	λ_b	3.15	0.91	(1.84,5.36)
Int. Censored	α_a	1.54	0.32	(1.01,2.28)
Int. Censored	α_b	3.97	1.06	(2.33,6.42)

Table 4.5. Posterior Probabilities for $\beta_2 < 0$

Data	$P(\beta_2 < 0)$
Lower	0.9951
Upper	0.9908
IC	0.9924

Table 4.6. Bayesian Estimates for Δ_λ

Method	Mean	S.D.	95% C.S.
Lower	3.57	1.32	(0.95,6.21)
CDC	3.99	1.56	(0.87,7.09)
IC	3.76	1.43	(0.92,6.63)

Table 4.7. Posterior Probabilities for Δ_λ Quantities

Data	$P(\Delta_\lambda > 3)$	$P(\Delta_\lambda > 4)$
Lower	0.6771	0.3593
Upper	0.7542	0.4960
IC	0.7098	0.4198

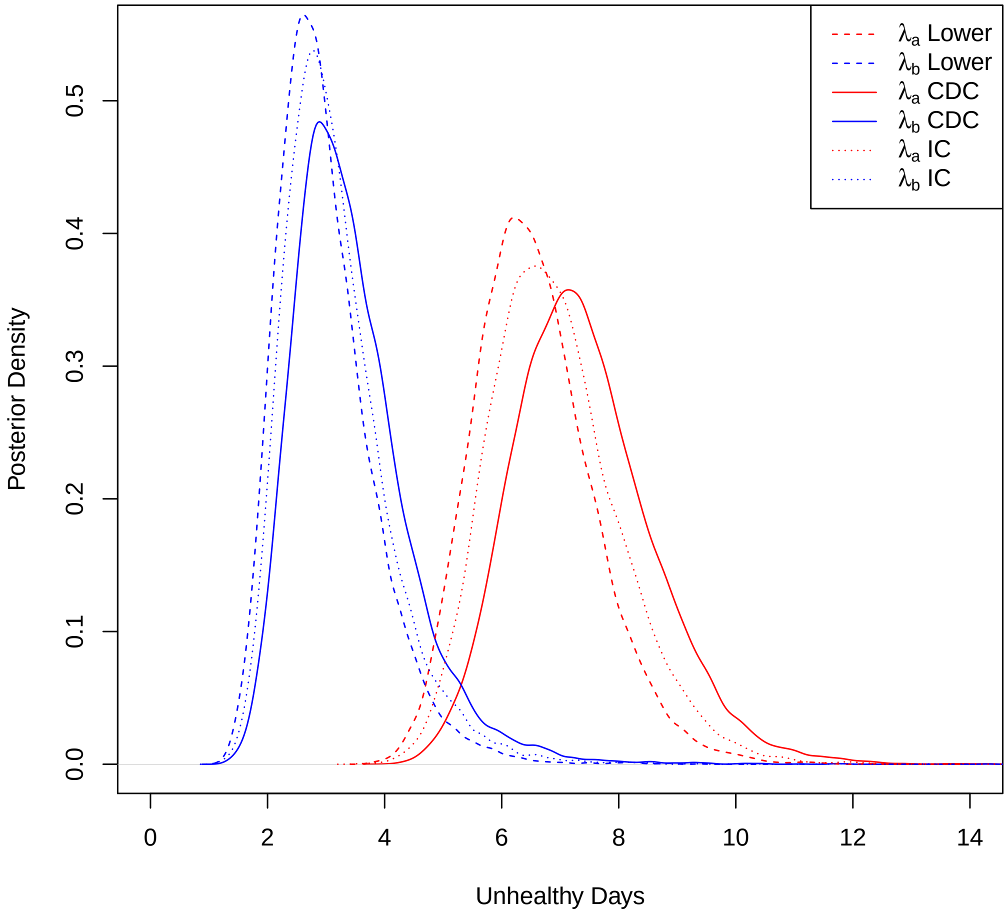

Figure 4.4: Posterior densities for λ_a (red) and λ_b (blue) based on the three data methods: Lower (dashed), CDC (solid), and interval-censored (dotted).

CHAPTER FIVE

Conclusions

In this dissertation we developed a model for accomodating interval-censored count data that are overdispersed relative to the Poisson distribution. We developed the interval-censored negative binomial likelihood for handling this type of data, and also explored the likelihood model for regressing the interval count responses on a single binary covariate. We presented two distinct survey examples in which this type of model might be useful, and demonstrated frequentist and Bayesian analyses under various considerations.

In Chapter 2 we defined the interval-censored negative binomial likelihood. We presented examples of factors that have an effect on the joint likelihood for the dispersion and expected count parameters, including interval-censoring widths, percent of the data that is censored, sample size, and the positioning of the precise count within the interval. In general, there is more differentiation between the precise and interval-censored cases for increased interval widths or percentage of interval-censored data, and for decreased sample size. If the precise counts are centered within the intervals, the shapes of the precise and interval-censored likelihood contours are very similar, but as the precise counts tend toward the lower or upper endpoints of the intervals, the interval-censored likelihood may have a drastically different shape and location from that of the precise.

For the frequentist approach to estimation of the likelihood, we detailed the Newton-Raphson iterative method for finding the MLEs and associated standard errors for the parameters, and 95% Wald approximate interval estimates. We illustrated that, especially for heavy interval censoring, the distributions of the MLEs may not always be well approximated by the normal distribution, and thus the Wald intervals may be problematic. For the Bayesian approach to the problem, we investigated several prior distribution options for the dispersion parameter, r. We found the frequently used gamma(0.001,0.001) prior to be problematic in some cases, and we investigated several other gamma priors with various negative binomial likelihoods in a prior-to-posterior sensitivity study. We did not find a general choice of gamma prior that would be relatively noninformative in the posterior across our likelihood examples. We also suggested a uniform(0,1000) prior for r, which provided an alternative to the gamma prior structure but is also not indicated for general use, primarily because of MCMC convergence issues. The latter can be avoided by using a less diffuse prior on r.

Lastly in Chapter 2, we presented a simulation study to further explore some of the behaviors observed in the examples. We began with precise negative binomial data and applied interval-censoring to the precise counts, then compared four different methods of handling the interval-censored data in terms of their ability to estimate the true values of the parameters. For a large expected count and a relatively small dispersion parameter, we found that the interval-censored and mean methods were superior to the lower and upper methods, based on simulation average bias and interval estimate coverage. There was little differentiation between the interval-censored and mean methods for these data sets. However, for a data scenario with small expected count and relatively large dispersion parameter, we demonstrated that the interval-censored method out-performed the mean method as well as the lower and upper methods.

In Chapter 3, we introduced the interval-censored negative binomial regression model. In our examples utilizing this model we focused on the case of a single binary regression covariate. We outlined an example from the General Social Survey (GSS) for which interval-censored responses might be plausible, and we ran frequentist and Bayesian analyses on a smaller sample that was generated based on parameter estimates from the original GSS data. Due to the diffuse prior employed in the Bayesian analysis, the parameter estimates based on the two methods were very similar. Just as in Chapter 2, we investigated the appropriateness of the Wald interval approximations and found that the normal approximations to the distributions of the MLEs did not always provide a good fit—thus, the Wald intervals would not provide good estimates. In the Bayesian context we also discussed several choices of prior for the regression coefficients. These choices were a Cauchy(0,2.5), a traditional diffuse normal, and a conditional means prior based on past data or expert opinion. We found that, for our small GSS data example, the Cauchy and conditional means priors created slightly more differentiation between the interval-censored and precise marginal posteriors than was exhibited using the normal priors.

We presented examples of likelihood contours for the parameters based on varying interval-censored widths, percentage of interval-censored data, sample size, positioning of the precise counts in the intervals, and fixed values of the dispersion parameters. The trends in the shapes and locations of the contours were similar to what we observed in our likelihood examples in Chapter 2. Finally in Chapter 3, we presented a simulation study using parameter values similar to those estimated in the GSS data set. In addition to the bias, interval widths, and interval coverages used for comparisons in the Chapter 2 simulation study, here we also computed the significance of the regression parameter on the single binary covariate, based on the percentage of interval estimates for this parameter that did not contain zero. In our first data scenario we found that the interval-censored method performed similarly to the mean method for estimates on two of the parameters, but the interval-censored method was superior to the mean method, in terms of bias and interval coverage of the true parameter, for three of the parameters. The lower method appeared to have reduced bias and better coverage in estimating the two dispersion parameters, but at a cost of much wider credible sets. When the underlying true values of the expected counts were altered to be closer together, we saw very similar results with an expected decrease in significance of the regression parameter on the covariate.

We concluded with an applied chapter, in which we used a subset of actual data from the 2003 California Health Interview Survey (CHIS) and compared the interval-censored negative binomial regression model on this data to two precise-data alternatives (equivalent to the lower and upper methods detailed in our simulation studies). We found that the interval-censored model provided estimates that took on an intermediate value between those of the lower and upper methods. Though there was not much differentiation among the three data methods due to the small percentage of interval-censoring in the data, the interval-censored estimates would be preferred due to the retention by this method of all information provided by the survey respondents.

In future research, one of the issues we would like to address in the frequentist approach is the inadequate normal approximation to the sampling distributions of the MLEs in some cases that leads to unreliable Wald interval estimates. We will investigate higher-order approximations to these sampling distributions that could yield better interval estimates for the parameters, including Barndorff-Nielsen's (1983) $p^\star$ approximation. We will also introduce additional covariates—both binary and continuous—into the regression model, and develop further simulation studies based on additional covariate values.

APPENDICES

APPENDIX A

Convergence Plots for First Example from Section 2.6

In Section 2.6 we ran several Bayesian analyses in *WinBUGS*. For our first example we used generated data of $n = 10$ exact counts from a $NB(10, 0.1)$ distribution to form $\mathbf{y}$, and a corresponding interval-censored data in which $w_i = 1 \; \forall \; i$. In *WinBUGS* we ran two chains with 5000 burn-in samples, followed by 150000 updates which reduced to chain lengths of 5000 each after thinning of 30. Convergence diagnostics for the precise and interval-censored analyses are plotted in Figures A.1 and A.2, respectively.

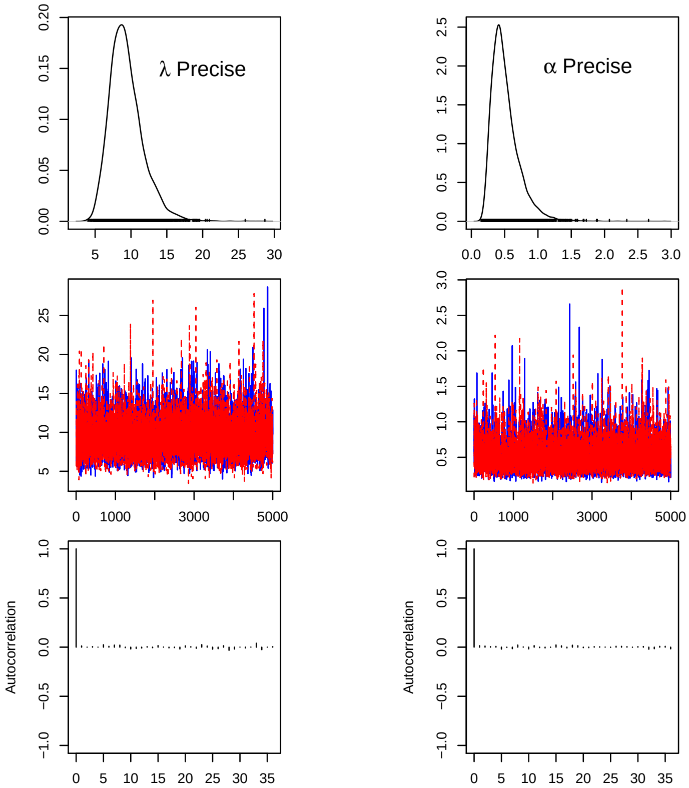

Figure A.1: Convergence diagnostics for a *WinBUGS* run from Section 2.6, precise case. Convergence is plotted for λ in the left column, and α in the right column: smoothed posterior densities (top row), trace plots (center row), and autocorrelation plots (bottom row).

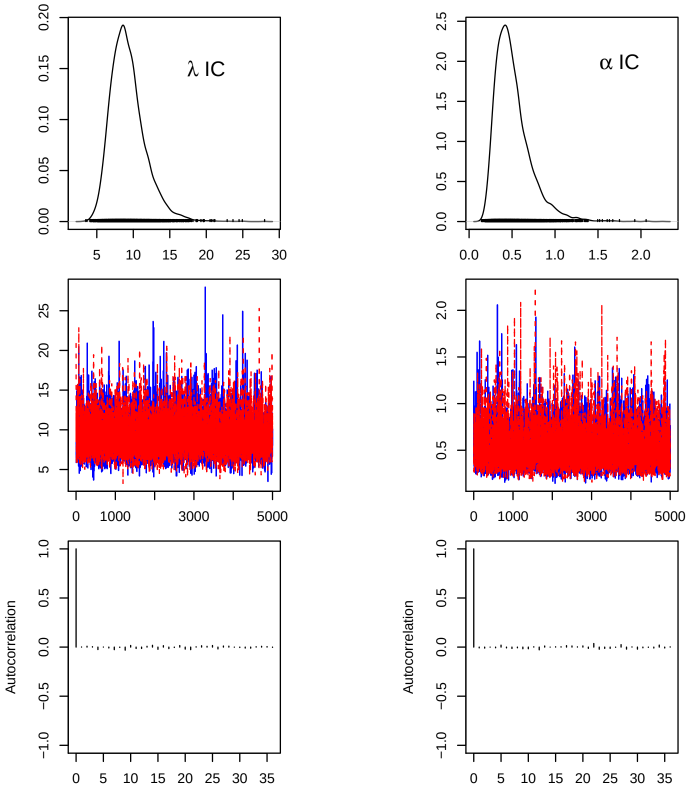

Figure A.2: Convergence diagnostics for a *WinBUGS* run from Section 2.6, interval-censored case. Convergence is plotted for λ in the left column, and α in the right column: smoothed posterior densities (top row), trace plots (center row), and autocorrelation plots (bottom row).

APPENDIX B

High Performance Computing Specification

Baylor University provides several high-performance computer systems for academic and research computing resources. For the simulations in Section 2.7 we use the Kodiak system. Kodiak has 128 nodes with dual-quadcore processors, for a total of 1024 cores available for use. Multiple users may utilize the system for a number of computing jobs through the batch processing system. Each job has its own dedicated core for the duration of the job. Each node has 16 GB of memory available which is shared across the 8 cores. Specifications for the Kodiak system are given in Table B.1.

Table B.1. Kodiak Cluster Specifications

Model	HP C3000BL
Operating System	CentOS
Kernel	2.6.9-67.0.4.EL-SFS2.3_0smp
Cluster Mgt. Software	Platform Manage
Hardware	Dual quadcore Intel Xeon 5355, 16 GB RAM

APPENDIX C

Simulation Results from Section 2.7

In Section 2.7 we presented an interesting subset of our simulation results. In particular, we focused on the case of $\lambda = 30$, $\alpha = 0.05$, and interval-censored widths of $w_i = 10 \; \forall \; i$. For the $\lambda = 30$ scenario, we also ran the simulation with $w_i = 1 \; \forall \; i$ and, separately, $w_i = 5 \; \forall \; i$. The precise, interval-censored, mean, lower, and upper methods for these two cases behaved similarly as in the $w_i = 10$ case, with the notable difference being the reduction in bias and greater 95% interval coverage for the lower and upper methods. We present the 95% interval results under the $w_i = 5$ case for λ in Figure C.1 and Table C.1, and results for α in Figure C.2 and Table C.2.

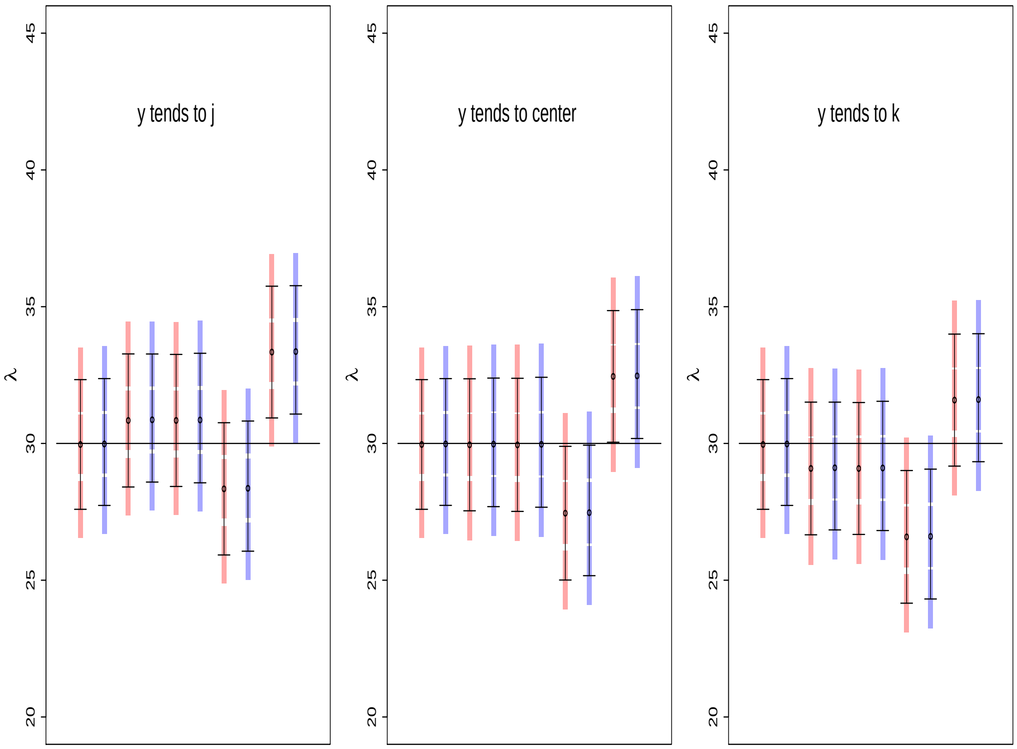

Figure C.1: Frequentist (red) and Bayesian (blue) simulation results for λ, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 30$ and $\alpha = 0.05$, and interval-censored widths are all equal to 5. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

Table C.1. Coverage for λ where true $\lambda = 30$ and IC widths=5

		Freq CI Widths			Bayes CS Widths		
y_i Tend	Method	Avg	SD	Covg	Avg	SD	Covg
j_i	Prec	4.74	0.48	0.96	4.64	0.50	0.96
j_i	IC	4.87	0.51	0.92	4.68	0.48	0.90
j_i	Mean	4.83	0.46	0.92	4.74	0.49	0.91
j_i	Lower	4.83	0.48	0.78	4.76	0.49	0.78
j_i	Upper	4.82	0.46	0.16	4.69	0.48	0.13
None	Prec	4.74	0.48	0.96	4.63	0.50	0.96
None	IC	4.83	0.51	0.97	4.70	0.48	0.94
None	Mean	4.87	0.46	0.94	4.75	0.48	0.94
None	Lower	4.89	0.47	0.44	4.77	0.48	0.46
None	Upper	4.81	0.45	0.48	4.71	0.47	0.47
k_i	Prec	4.74	0.48	0.96	4.64	0.50	0.96
k_i	IC	4.85	0.75	0.88	4.67	0.54	0.89
k_i	Mean	4.82	0.52	0.89	4.73	0.54	0.89
k_i	Lower	4.85	0.55	0.22	4.75	0.54	0.22
k_i	Upper	4.83	0.54	0.77	4.68	0.53	0.69

Table C.2. Coverage for α where true $\alpha = 0.05$ and IC widths=5

		Freq CI Widths			Bayes CS Widths		
y_i Tend	Method	Avg	SD	Covg	Avg	SD	Covg
j_i	Prec	0.07	0.01	0.92	0.06	0.01	0.92
j_i	IC	0.06	0.01	0.90	0.06	0.01	0.87
j_i	Mean	0.06	0.01	0.92	0.06	0.01	0.91
j_i	Lower	0.08	0.01	0.97	0.07	0.01	0.94
j_i	Upper	0.05	0.01	0.84	0.05	0.01	0.83
None	Prec	0.07	0.01	0.92	0.06	0.01	0.92
None	IC	0.07	0.01	0.94	0.06	0.01	0.93
None	Mean	0.07	0.01	0.96	0.07	0.01	0.94
None	Lower	0.08	0.01	0.97	0.08	0.01	0.93
None	Upper	0.06	0.01	0.87	0.05	0.01	0.85
k_i	Prec	0.07	0.01	0.92	0.06	0.01	0.92
k_i	IC	0.07	0.01	0.93	0.07	0.02	0.93
k_i	Mean	0.07	0.01	0.94	0.07	0.01	0.92
k_i	Lower	0.09	0.02	0.95	0.08	0.02	0.87
k_i	Upper	0.06	0.01	0.87	0.06	0.01	0.86

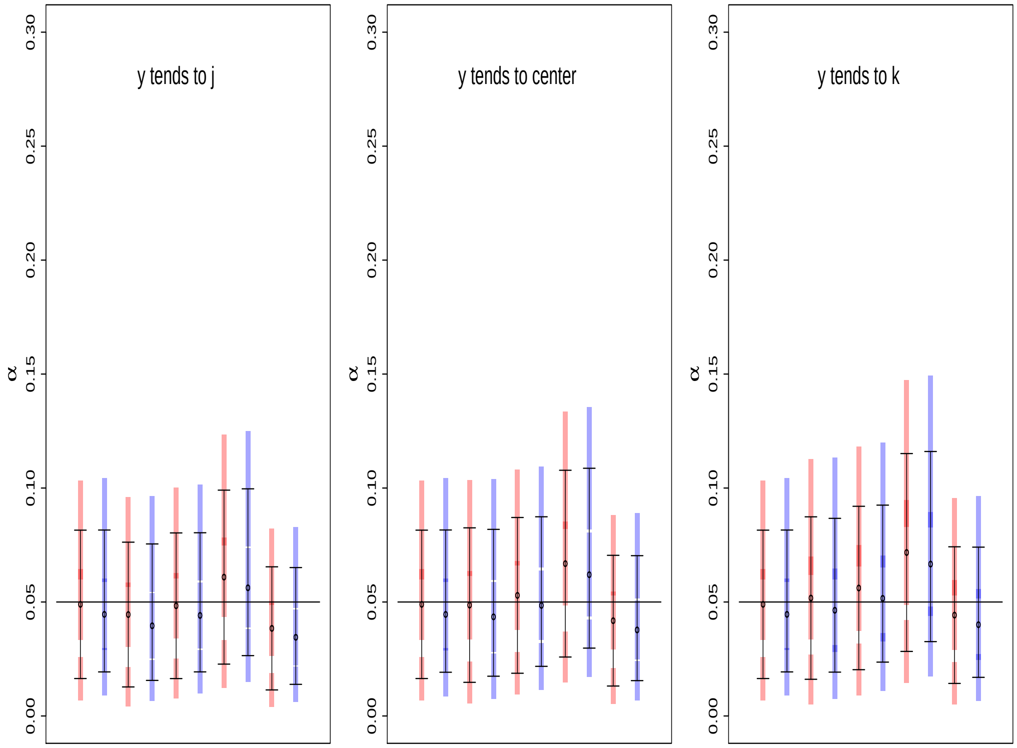

Figure C.2: Frequentist (red) and Bayesian (blue) simulation results for α, based on three different tendencies for y_i in $[j_i, k_i]$, where originating $\lambda = 30$ and $\alpha = 0.05$, and interval-censored widths are all equal to 5. Within each panel, the methods are: precise, interval-censored, mean, lower, upper.

APPENDIX D

Additional Simulation Results from Section 3.7

In Section 3.7 we presented results for simulations based on the true underlying expected counts $\lambda_1 = 7$ and $\lambda_1 = 3$. Here we give the full results for our alternative cases with $\lambda_1 = 7$ and $\lambda_2 = 6$. (Refer to Table 3.16 for Case 3 and Case 4 results on β_2.)

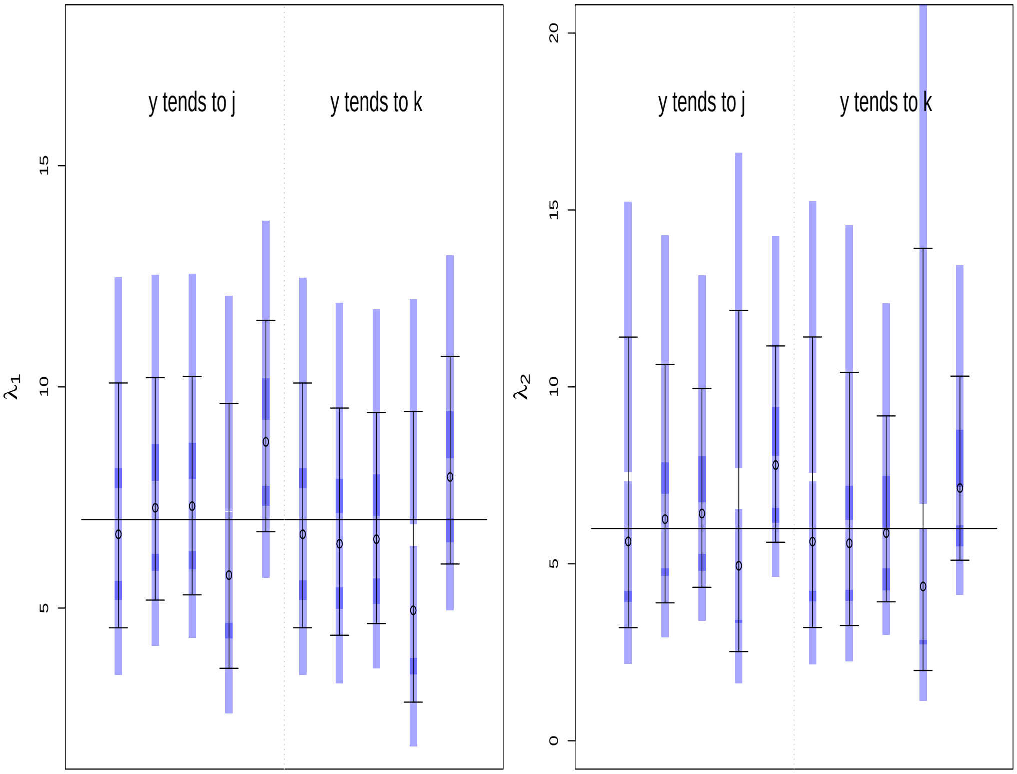

Figure D.1: Bayesian simulation results for λ_1 (left panel) and λ_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$ and $\lambda_2 = 6$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

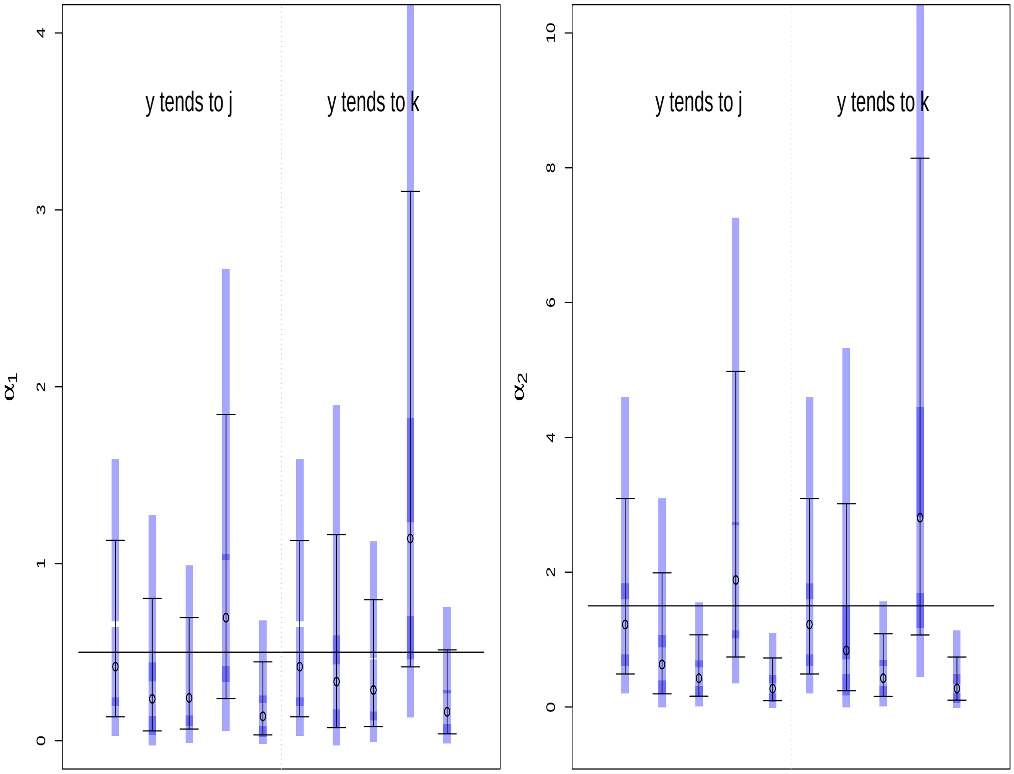

Figure D.2: Bayesian simulation results for α_1 (left panel) and α_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$, $\lambda_2 = 6$, $\alpha_1 = 0.5$ and $\alpha_2 = 1.5$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

Table D.1. Coverage for $\lambda_1(= 7)$ and $\lambda_2(= 6)$

			y_i tends to j_i				y_i tends to k_i		
Param	Method	Est	Avg	SD	Cov	Est	Avg	SD	Cov
λ_1	Prec	6.67	5.53	1.64	0.88	6.67	5.53	1.64	0.88
λ_1	IC	7.27	5.03	1.67	0.88	6.45	5.14	1.69	0.82
λ_1	Mean	7.31	4.94	1.56	0.88	6.56	4.77	1.52	0.82
λ_1	Lower	5.75	5.99	1.82	0.82	4.95	6.57	2.03	0.78
λ_1	Upper	8.76	4.78	1.48	0.62	7.96	4.69	1.45	0.80
λ_2	Prec	5.63	8.21	3.24	0.42	5.63	8.21	3.26	0.42
λ_2	IC	6.26	6.74	3.09	0.22	5.58	7.15	3.75	0.42
λ_2	Mean	6.42	5.62	2.43	0.04	5.87	5.25	2.36	0.20
λ_2	Lower	4.95	9.64	4.02	0.70	4.36	11.93	7.05	0.88
λ_2	Upper	7.79	5.55	2.32	0.00	7.14	5.20	2.31	0.00

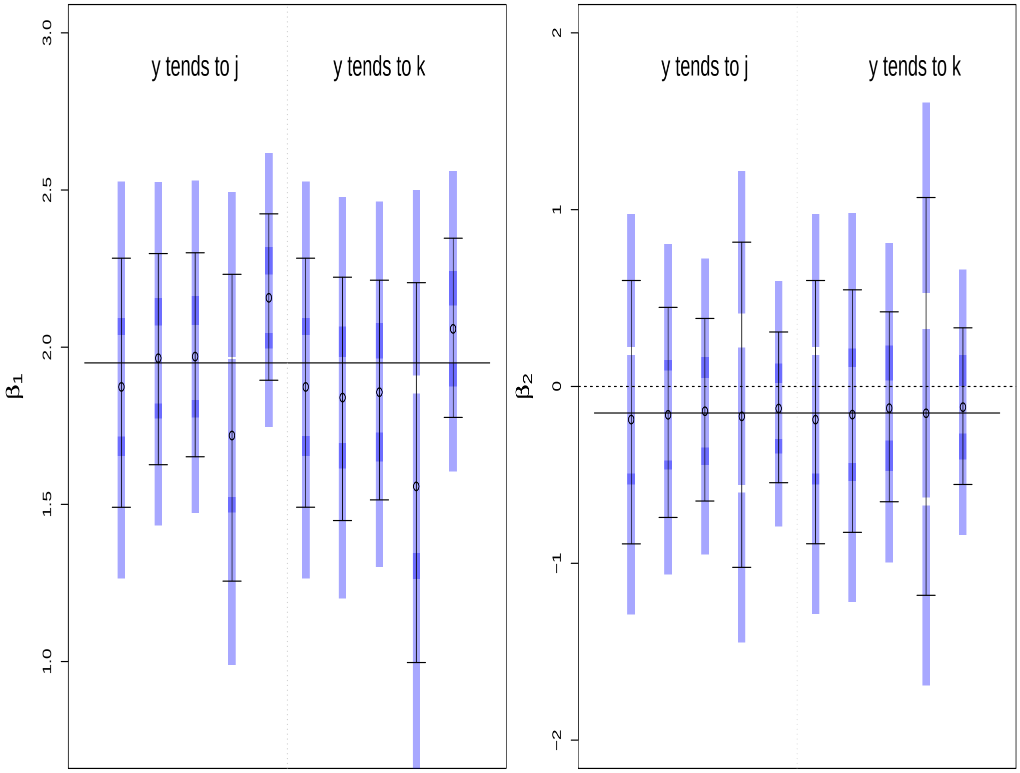

Figure D.3: Bayesian simulation results for β_1 (left panel) and β_2 (right panel), based on y_i tending toward j_i (left side) or k_i (right side), where originating $\lambda_1 = 7$, $\lambda_2 = 6$, $\beta_1 = 1.95$ and $\beta_2 = -0.15$. Within each panel section, the methods represented (from left to right) are: precise, interval-censored, mean, lower, upper.

Table D.2. Coverage for $\alpha_1(= 0.5)$ and $\alpha_2(= 1.5)$

			y_i tends to j_i				y_i tends to k_i		
Param	Method	Est	Avg	SD	Cov	Est	Avg	SD	Cov
α_1	Prec	0.42	1.00	0.37	0.90	0.42	1.00	0.37	0.90
α_1	IC	0.24	0.75	0.40	0.74	0.33	1.09	0.67	0.84
α_1	Mean	0.24	0.63	0.23	0.76	0.29	0.72	0.26	0.82
α_1	Lower	0.70	1.61	0.66	0.90	1.14	2.69	1.61	0.68
α_1	Upper	0.14	0.41	0.19	0.40	0.16	0.47	0.20	0.56
α_2	Prec	1.22	2.61	1.24	0.90	1.22	2.60	1.24	0.92
α_2	IC	0.63	1.79	0.95	0.62	0.84	2.77	2.12	0.76
α_2	Mean	0.43	0.91	0.34	0.16	0.43	0.93	0.34	0.22
α_2	Lower	1.89	4.24	1.94	0.94	2.81	7.07	4.75	0.80
α_2	Upper	0.27	0.63	0.28	0.02	0.28	0.64	0.30	0.04

Table D.3. Coverage for $\beta_1(= 1.95)$

	y_i tends to j_i				y_i tends to k_i			
Method	Est	Avg	SD	Cov	Est	Avg	SD	Cov
Prec	1.87	0.79	0.15	0.86	1.87	0.79	0.15	0.86
IC	1.97	0.67	0.15	0.88	1.84	0.77	0.18	0.82
Mean	1.97	0.65	0.12	0.88	1.86	0.70	0.13	0.82
Lower	1.72	0.98	0.19	0.82	1.56	1.21	0.28	0.78
Upper	2.16	0.53	0.10	0.62	2.06	0.57	0.10	0.80

APPENDIX E

Investigation of Uniform Prior on Dispersion Parameters, CHIS Data Subset

We look at the posterior means and standard deviations of β_1, β_2, α_a, α_b, λ_a, and λ_b in Figures E.1 and E.2. These are based on the Cauchy$(0, 2.5)$ priors for β_1 and β_2, and $r_j \sim \text{Uniform}(0, B)$ for $j = 1, 2$, with increasing values of B on the horizontal axis.

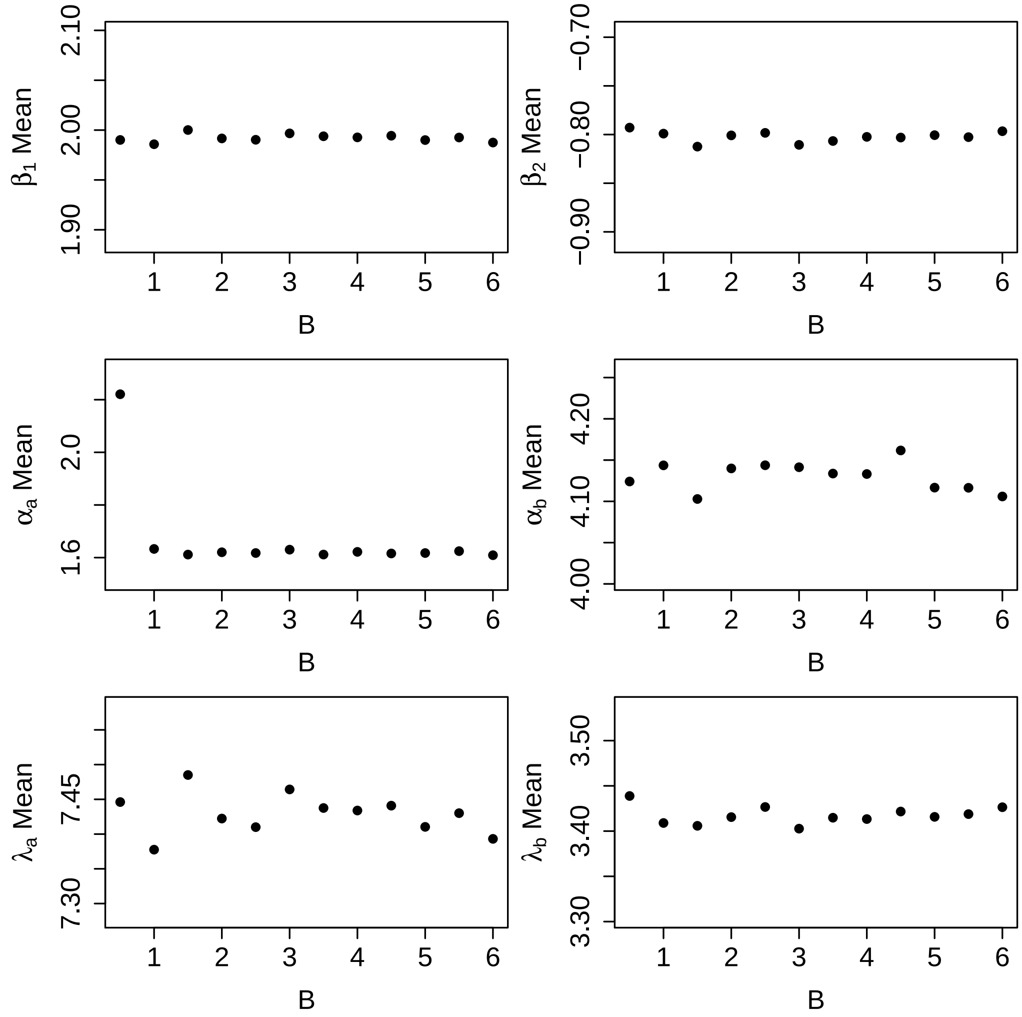

Figure E.1: Posterior means for β_1 (top left), β_2 (top right), α_a (center left), α_b (center right), λ_a (bottom left), and λ_b (bottom right), calculated for increasing values of B for $r_j \sim \text{uniform}(0, B)$, $j = 1, 2$.

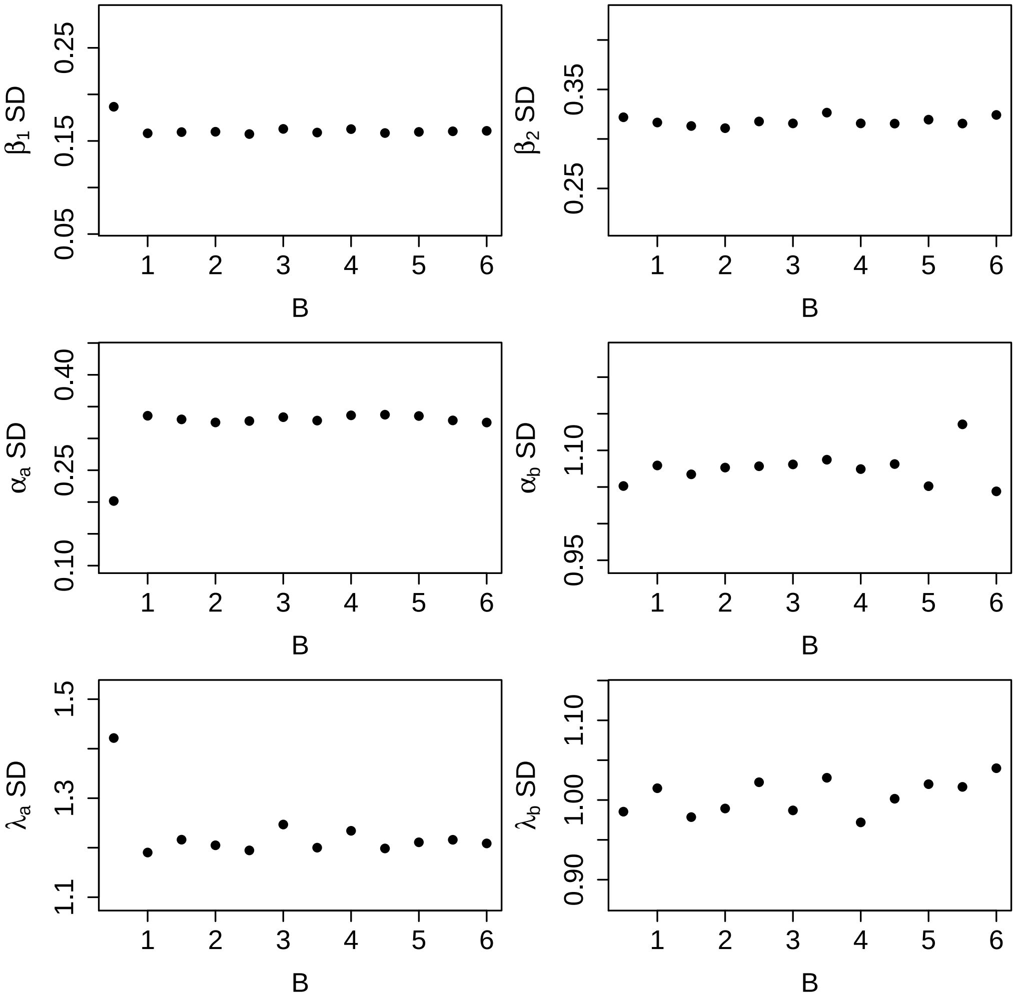

Figure E.2: Posterior standard deviations for β_1 (top left), β_2 (top right), α_a (center left), α_b (center right), λ_a (bottom left), and λ_b (bottom right), calculated for increasing values of B for $r_j \sim \text{uniform}(0, B)$, $j = 1, 2$.

The posterior means and standard deviations stabilize quickly—all reach relatively stable values by $B = 1$. Beyond $B = 6$ the changes in posterior means and standard deviations are minor; for our Bayesian analysis in Chapter 4 we select $B = 6$ so that the priors are $r_1 \sim$ Uniform(0,6) and $r_2 \sim$ Uniform$(0, 6)$. This allows minimum values for α_a and α_b of $1/6 = 0.17$; the lower 95% confidence limits for the α's given in Table 4.3 all exceed this value.

BIBLIOGRAPHY

Barndorff-Nielsen, O. and Cox, D. (1983), "On a formula for the distribution of the maximum likelihood estimator," *Biometrika*, 70, 343–365.

Bedrick, E., Christensen, R., and Johnson, W. (1996), "A New Perspective on Priors for Generalized Linear Models," *Journal of the American Statistical Association*, 91, 1450–1460.

Centers for Disease Control and Prevention (2000), "Measuring Healthy Days," www.cdc.gov/hrqol/pdfs/mhd.pdf.

Davies, S. and the R Core Team (2012), "Fitting Generalized Linear Models, R Documentation," stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html.

Gelman, A., Jakulin, A., Pittau, M., and Su, Y. (2008), "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models," *The Annals of Applied Statistics*, 2, 1360–1383.

Gentle, J. (2009), *Computational Statistics*, New York: Springer-Verlag.

Hilbe, J. (2011), *Negative Binomial Regression*, Cambridge, UK: University Press.

Hilbe, J. and Robinson, A. (2012), "Package 'COUNT'," cran.r-project.org/web/packages/COUNT/COUNT.pdf.

Khuri, A. (2003), *Advanced Calculus with Applications in Statistics*, New York: Wiley, 2nd ed.

Kobau, R., Zahran, H., Grant, D., Thurman, D., Price, P., and Zack, M. (2007), "Prevalence of Active Epilepsy and Health-related Quality of Life among Adults with Self-reported Epilepsy in California: California Health Interview Survey, 2003," *Epilepsia*, 48, 1904–1913.

National Opinion Research Center (2012), "About NORC," www3.norc.org/GSS+Website/About+GSS/About+NORC/.

Nelder, J. and Mead, R. (1965), "'A simplex algorithm for function minimization," *Computer Journal*, 7, 308–313.

Ntzoufras, I. (2009), *Bayesian Modeling using WinBUGS*, Hoboken, NJ: John Wiley & Sons.

Pawitan, Y. (2001), *In All Likelihood*, New York, NY: Oxford UP.

Pruszynski, J. (2010), "Bayesian Models for Discrete Censored Sampling and Dose Finding," Ph.D. thesis, Baylor University.

Spiegelhalter, D., Abrams, K., and Myles, J. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, New York: Wiley.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003), "WinBUGS User Manual," www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf.

The R Core Team (2012), "Fit a Negative Binomial Generalized Linear Model, R Documentation," stat.ethz.ch/R-manual/R-patched/library/MASS/html/glm.nb.html.

Watson, S. (2011), "Poisson Regression Models for Interval Censored Count Data," Ph.D. thesis, Baylor University.