Master Thesis

Bayesian Hierarchical Modelling of Zero-Inflated Faecal Egg Counts

Author(s):
Wang, Craig

Publication Date:
2015

Permanent Link:

https://doi.org/10.3929/ethz-a-010695750 →

ETH
Swiss Federal Institute of Technology Zurich

Seminar for Statistics

Department of Mathematics

Master Thesis Autumn 2015

Craig Wang

Bayesian Hierarchical Modelling of Zero-Inflated Faecal Egg Counts

Submission Date: April 10th 2016

Main Reader: Prof. Dr. Reinhard Furrer
Adviser: Prof. Dr. Reinhard Furrer

Preface

I would like to thank my supervisor Prof. Dr. Reinhard Furrer for all the support and guidance he offered throughout my master thesis. It has been a great pleasure to work with him. In addition, I would like to thank my parents and all my friends especially my girlfriend for supporting me during my master degree.

Craig Wang

10.04.2016 Zürich

Abstract

In this thesis, we investigate the statistical modelling of zero-inflated faecal egg counts to assess the efficacy of anthelmintic treatment.

The parasite infection in livestock is commonly treated with anthelmintic drugs, however the prevalence of anthelmintic resistance in parasites has increased over the past years. The reduction in faecal egg counts is a standard measure that is used to quantify the efficacy of anthelmintic drugs. In this thesis, we propose zero-inflated Bayesian hierarchical models to estimate the reduction. The models capture the occurrence of zero counts from unexposed livestock, the biological aggregation of parasites, as well as the variability in the data collection process.

A simulation study shows the proposed Bayesian models are more robust compared with other approaches. In addition, the credible intervals from Bayesian models offer better coverage of the true reduction parameter overall. A case study on Swedish sheep flocks further illustrates the advantages of the proposed model, it can provide a confidence interval for the estimated reduction in the cases where the conventional methods cannot. In comparison with the Bayesian model without zero inflation, the zero-inflated Bayesian model accounts for the potential unexposed animals in the population and captures the paired structure more accurately.

Contents

Chapter 1

Introduction

This chapter provides an introduction to the thesis. Section 1.1 outlines the importance of studying parasite infection, explains the life cycle of the parasites and introduces the diagnostic procedure. Section 1.2 introduces the statistical methods that can be used to analyze faecal egg count data. Section 1.3 highlights the contribution of this thesis with reference to the previous works. Finally, Section 1.4 details the structure for the rest of this thesis.

1.1 Epidemiology of Gastrointestinal Nematodes

Gastrointestinal nematodes (GINs) are parasitic worms that survive in livestock hosts, such as sheep, cattle and horses. The infection of GINs can lead to a series of problems including reduction in skeletal growth, live-weight gain and milk yield (Houtert and Sykes, 1996). According to the Food and Agriculture Organization (FAO, 2015), there were approximately 5 billion livestock in 2013, together they produced of 700 million tons of milk and 300 million tons of meat. Such infection, which is already very common in sheep (Mortensen et al., 2003; Zanzani et al., 2014), cattle (Waruiru et al., 2001; Pfukenyi et al., 2007; Tariq, 2014) and horses (Slocombe et al., 2008; Donato et al., 2009), can cause great economic burden on ruminant production in some regions (Waller, 1997; Perry and Randolph, 1999).

There are numerous species of GINs, such as *Haemoncus contortus*, *Strongylus vulgaris* and *Nematodirus battus* etc., all of which share a similar life cycle, as illustrated in Figure 1.1. During the grazing season, the animals ingest larvae that survive in the herbage. Then larvae mature inside animals' body to develop into adult worms with the ability of sexual reproduction. Through further growth, the adult worms will lay eggs in the intestine of infected animals, where the eggs are then shed in faeces and transported to the herbage. Finally, the larvae eggs begin to develop and wait to be ingested by the animals again.

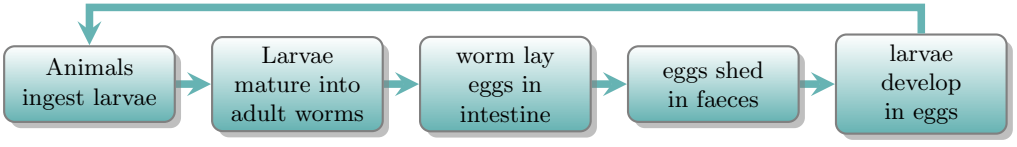


Figure 1.1: A flow chart of a typical GINs' life cycle.

The regular administration of anthelmintic treatments is a widely used method to control the infection. It aims to reduce the infection level and to prevent its transmission, hence reduces production losses. However, anthelmintic resistant nematodes appeared in different regions across the globe since late 1950s (Kaplan, 2004). The extensive use of anthelmintic treatments has led to an increasing problem of anthelmintic resistance. Once resistance is detected, alternative anthelmintic treatments are needed in order to avoid further production losses. Accurate and reliable methods to assess the treatment efficacy and resistance status are thus essential to effectively monitor and control the infections.

A standard measure for the efficacy of anthelmintic treatment is the reduction in the number of parasite eggs in animal faeces. The World Association for the Advancement of Veterinary Parasitology (WAAVP) guidelines (Coles et al., 1992) detailed a modified McMaster counting technique to measure the faecal egg counts (FECs) in animal faeces. The basic steps of the McMaster technique are as follows:

1. A faecal sample is diluted with flotation fluid and then soaked;
2. The solution is mixed thoroughly to ensure homogeneous distribution of the eggs;
3. Then the solution is filtered with a mesh to remove large debris;
4. After some further processing, a small sample of the resulting solution is placed into the chambers of a McMaster counting slide;
5. Finally, the number of eggs inside the grid area of the chamber is counted under microscope.

The number of eggs per gram of faeces (epg) is calculated, by multiplying the observed count with the dilution factor and then divide by the volume of the McMaster chamber. The default unit for reporting FECs is epg and it is also the unit used for calculating the reduction in FECs.

1.2 Statistical Modelling of Faecal Egg Counts

1.2.1 Classical Approaches

The conventional method to estimate the FECs in a livestock population is recommended by Coles et al. (1992), which simply uses the arithmetic mean across the samples. However, this method ignores variabilities introduced by the McMaster counting technique and also by the different infection levels across animals. The same authors recommended to calculate the reduction in FECs in a similar manner, by computing the ratio of arithmetic means for pre-treatment and post-treatment FECs. The point estimate is equivalent to the result from a Poisson regression (see Section 3.1 for the proof), however its interval estimate is often unreliable. It has been pointed out that there are considerable variability in the estimation of FECs reduction using this ratio (Miller et al., 2006).

Parametric approaches have gained popularity in recent years. The distribution of parasites and egg counts are known to be overdispersed within the host population (Grenfell et al., 1995), and such counts are often well described by negative binomial distributions. Torgerson et al. (2005) assumed a negative binomial distribution on the FECs and applied a parametric bootstrap method. Negative binomial regression can also be applied to model the FECs. In order to quantify the reduction, the log baseline counts is used as an offset

term (Alexander Neal, 2012). Pullan et al. (2010) applied the negative binomial regression in a Bayesian setting. Nødtvedt Ane et al. (2002) used a multivariate zero-inflated negative binomial (ZINB) regression to investigate the level of infection across different regions and age groups of Canadian dairy cows. Denwood et al. (2008) later showed the distribution of FECs of a particular nematode *N.battus* follows a ZINB distribution, where the zero-inflation comes from the unexposed animals in an infected population. However in the analysis of FECs reduction, the sample size is often too small such that the ZINB regression fail to give sensible results.

1.2.2 Bayesian Hierarchical Models

An alternative class of methods is Bayesian hierarchical models. They are based on Bayes' theorem, using information from the data to update prior belief about the distribution of parameters, in order to obtain the posterior distribution about the parameters.

Bayes' theorem for continuous parameters can be written as

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}, \tag{1.2.1}$$

where θ is a vector of the parameters, and D denotes the data. $p(\theta|D)$ is the posterior distribution that describes the parameters, based the data and taking the prior information into account. $p(\theta)$ is the prior distribution of the parameters, which can be specified using common probability density functions such as a Normal or a Gamma distribution. The prior distribution generally contains very little information comparing to the data. $p(D|\theta)$ is the likelihood, which also takes the form of a density distribution, describes the likelihood of data conditional on the parameters θ according to some pre-specified distribution. Finally, $p(D)$ is the prior probability of data. One may consider the data as fixed such that $p(D)$ is a constant, hence it is also very common to drop the term $p(D)$ and write Eq (1.2.1) as $p(\theta|D) \propto p(\theta)p(D|\theta)$.

Like other parametric methods, Bayesian hierarchical models are able to capture the extra variabilities that exist in the data by specifying certain distribution for the likelihood functions. In addition, the likelihood can be specified at different hierarchical levels to allow for different layers of variability. In this case, one layer can be for the McMaster counting variability and another layer can be for the overdispersion. It also allows for incorporating prior belief about the FECs, such as the information about distribution parameters from previous study results. Recently, Denwood et al. (2010); Paul et al. (2014) and Levecke et al. (2015) proposed different Bayesian models in effort to improve the model for estimating the reduction in FECs.

Markov chain Monte Carlo (MCMC) methods are often used in conjunction with Bayesian hierarchical models. They simulate draws of random numbers approximately from the posterior distributions. Two commonly used algorithms are the Gibbs sampler (George Casella, 1992) and the Metropolis-Hastings (MH) algorithm (Hastings, 1970). The implementation of Bayesian hierarchical models using those algorithms are often complex if they are coded manually. They require case-by-case implementation of prior distributions, likelihood specifications, proposal distributions and tuning parameters. For a multi-level hierarchical model, it can lead to extensive codes and hence hard to maintain.

Stan (Carpenter, 2015) is a open-sourced modelling language for Bayesian inference, it can be easily utilized via the `rstan` package (Guo et al., 2015) in R (R Core Team, 2015).

The Stan programs written in R are flexible in its hierarchical modelling structure and intuitive in its syntax (Appendix A), the programs are compiled into C++ code for efficient computation upon calling. Stan employs Hamiltonian Monte Carlo (HMC) (Homan and Gelman, 2014), it explores a multidimensional parameter space through its gradient and update all the parameters simultaneously. The derivative is computed analytically, which is much faster than numerical differentiation especially when the dimension of the model is large, hence it leads to dramatic improvement on the convergence speed compared to other inference tools such as BUGS (Lunn et al., 2000) and JAGS (Plummer, 2003). The posterior samples are less correlated, hence often a shorter MCMC chain is required until convergence. Previously, the usefulness of HMC relies on correctly tuning the two parameters, namely the step size and the desired number of steps. A step size being too large can cause low acceptance rates while being too small can lead to wasted computations. An inappropriate choice of the desired number of steps can result non-convergence. An enhanced version of the HMC is called the No-U-Turn sampler (NUTS) (Homan and Gelman, 2014). It eliminates the need to manually tune those two parameters, and hence making it easier to effectively use HMC. In this thesis, we utilize the Stan language via the `rstan` package to implement the Bayesian models.

1.3 Previous Works and Thesis Contributions

This thesis is based on the work of Paul et al. (2014) and the `eggCounts` package version 0.4-1 (Paul, 2015) available on CRAN. Paul et al. (2014) proposed a Bayesian hierarchical model using MH algorithm and Gibbs sampler to estimate the FECs reduction. An easy-to-use web interface was also implemented and made available online (Torgerson et al., 2014). The proposed model did not address an important feature of the data, which is the zero-inflation component. Although some additional models were subsequently implemented in the eggCounts package, the implementations are complex due to the dimensionality of the hierarchical models. In addition, several software bugs were discovered during the usage of the package, which occasionally resulted incorrect inferences on the reduction.

To address the aforementioned issues, this thesis contains the following contributions:

1. Extensively tested the existing `eggCounts` package on CRAN, and discovered some models resulted bias and poor coverage due to incorrect implementation;
2. Fixed some bugs regarding the data simulation process and error reporting in the `eggCounts` package;
3. Improved the performance of the Bayesian models proposed by Paul et al. (2014);
4. Proposed zero-inflated hierarchical models to estimate the reduction in FECs;
5. Conducted a comprehensive simulation study comparing various approaches to estimate the FECs reduction;
6. Discovered the relationship between the faecal egg count reduction test and the maximum likelihood estimator of the Poisson regression;
7. Showed the marginal distribution of our proposed model follows a ZINB distribution;
8. Re-wrote the `eggCounts` package (Appendix B) and implemented the models using

Stan, which allows for more flexible prior specifications and faster MCMC convergence; and

9. Updated the user-friendly web interface with new options including the proposed models (Appendix C).

1.4 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 contains the working paper *Zero-Inflated Hierarchical Models for Faecal Egg Counts to Assess Anthelmintic Efficacy*, which proposes and discusses the Bayesian hierarchical models with zero-inflation. Chapter 3 presents the derivation of some proofs. Finally, Chapter 4 summarizes this thesis and suggests future research directions.

Chapter 2

Working Paper: Zero-Inflated Hierarchical Models for Faecal Egg Counts to Assess Anthelmintic Efficacy

2.1 Abstract

The prevalence of anthelmintic resistance has increased in recent years, as a result of the extensive use of anthelmintic drugs to reduce the infection of parasitic worms in livestock. In order to detect the resistance, the number of parasite eggs in animal faeces is counted. Typically a subsample of the diluted faeces is examined, and the mean egg counts from both untreated and treated animals are compared. However, the conventional method ignores the variabilities introduced by the counting process and by different infection levels across animals. In addition, there can be extra zero counts, which arise as a result of the unexposed animals in an infected population. In this paper, we propose the zero-inflated Bayesian hierarchical models to estimate the reduction in faecal egg counts. The simulation study compares the Bayesian models with the conventional faecal egg count reduction test and other methods such as bootstrap and quasi-Poisson regression. The results show the Bayesian models are more robust and they perform well in terms of both the bias and the coverage. We further illustrate the advantages of our proposed model using a case study about the anthelmintic resistance in Swedish sheep flocks.

2.2 Introduction

Gastrointestinal nematodes are parasitic worms that survive in livestock hosts, such as sheep, cattle and horses. The infection is common in the livestock populations in some regions (Waruiru et al., 2001; Mortensen et al., 2003; Pfukenyi et al., 2007; Tariq, 2014; Zanzani et al., 2014). Such infection can lead to numerous problems including reduction in skeletal growth, live-weight gain and milk yield (Houtert and Sykes, 1996), which can impose great economic burden on ruminant production (Perry and Randolph, 1999). The regular administration of anthelmintic treatments is a widely used method to control the

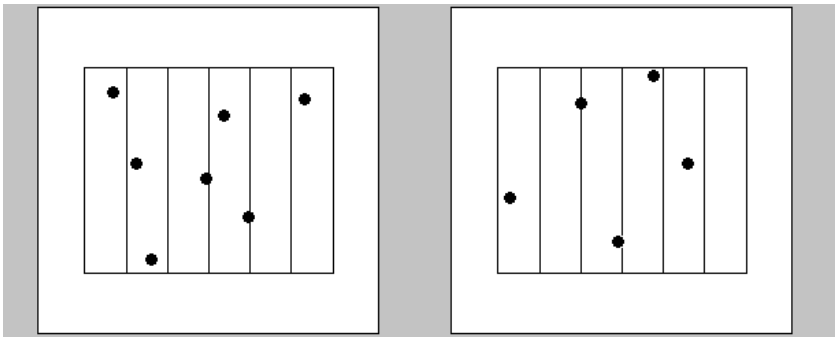


Figure 2.1: A schematic of a McMaster slide containing two chambers. There are 12 eggs (black dots) in total lying inside the grid area. Assuming an analytical sensitivity of 50, the number of eggs per gram of faeces is $12 \times 50 = 600$ epg.

infection. It aims not to eliminate the infection, but to reduce the infection intensity and prevent transmission (Levecke et al., 2012). However, anthelmintic resistant nematodes appeared in different regions across the globe since late 1950s (Kaplan, 2004). The extensive use of anthelmintic treatments has led to an increasing problem of anthelmintic resistance. Once a resistance is detected, alternative treatments are needed in order to avoid any further production losses. Accurate and reliable methods to assess the treatment efficacy are thus essential to effectively control and monitor the infection.

The widely used faecal egg count reduction test (FECRT) was established in the early 1990s (Coles et al., 1992). It is a straightforward method to calculate the reduction in faecal egg counts (FECs), by comparing the mean pre-treatment and post-treatment FECs. For sheep and goats, if both the percentage reduction in mean FECs is less than 95% and the corresponding lower confidence limit is less than 90%, then the anthelmintic resistance is declared to be present.

A standard method to obtain the FECs, the modified McMaster counting technique, is detailed in the guideline of the World Association for the Advancement of Veterinary Parasitology (WAAVP) (Coles et al., 1992). The basic procedures are as follows. Firstly, a faecal sample is diluted with water and soaked for some time, then it is mixed thoroughly to ensure homogeneous distribution of the parasite eggs. Afterward, it is filtered with a mesh to remove large debris. After some further processing, a small sample of the diluted solution is placed onto the chambers of a McMaster slide as shown in Figure 2.1. Finally, the number of eggs falls inside the grid of the slide is counted under microscope. The number of eggs per gram of faeces (epg) is calculated by multiplying the observed count with the dilution factor, and then divide by the volume of the McMaster chamber. The final multiplication factor is also known as the analytical sensitivity, which is typically 15 for high sensitivity and 50 for low sensitivity. Although this standard method was widely used in practice, some limitations have been pointed out in recent years.

First of all, the McMaster counting technique introduces substantial variability in the results which is not accounted for in the FECRT (Torgerson et al., 2012). As a consequence of this, the estimated efficacy were found to be quiet variable particularly for the samples with low pre-treatment FECs and efficacy in the range between 90%-95% (Miller et al., 2006). Secondly, the distribution of egg counts is typically aggregated or overdispersed within the host population (Grenfell et al., 1995). Levecke et al. (2012) evaluated the FECRT under different scenarios, highlighted the test results should be interpreted with

caution when the sample size is small and the aggregation level is high. There were several attempts to propose more elaborate statistical models in the past years. Dobson et al. (2012) proposed a novel way to determine the confidence limits of the FECs reduction, however it requires high counts and high analytical sensitivity. Torgerson et al. (2005) assumed a negative binomial distribution for the counts, and used parametric bootstrap to calculate the confidence interval (CI) of the FECs reduction. Denwood et al. (2010) considered a Poisson-gamma distribution for the counts, with the post-treatment mean linked to the pre-treatment mean via a scale factor. The inference is then done using Markov chain Monte Carlo (MCMC). Paul et al. (2014) proposed a hierarchical model that uses binomial distribution to capture the counting variability, and a Poisson-gamma distribution to model the overdispersion. The posterior median for the reduction and its 95% highest posterior density (HPD) interval is used for its point and interval estimate respectively. Levecke et al. (2015) proposed another Bayesian model with a slight different formulation. It used a Poisson distribution to capture the variability in the counting process and a negative binomial distribution to capture the overdispersion. To the best of our knowledge, a common assumption made by those recent Bayesian models is, all animals in an infected population are exposed. However, Denwood et al. (2008) showed the underlying distribution of the nematodes FECs can be zero-inflated negative binomial (ZINB). The zero-inflation component arises as a result of the unexposed livestock in an infected population.

In this paper, we propose the zero-inflated Bayesian hierarchical models to estimate the reduction in FECs. The models account for the counting variability and the between-animal variations. In addition, they consider the extra zero counts separately by introducing the zero-inflation components. The rest of this paper is organized as follows. Section 2.3 briefly reviews the conventional FECRT and efforts made to modify it. Section 2.4 introduces the zero-inflated Bayesian hierarchical models. Section 2.5 conducts a simulation study, where the bias and coverage of the estimated FECs reduction are compared across different methods. In Section 2.6, a case study is used to illustrate the proposed methods for estimating the reduction in FECs, where the anthelmintic resistance in a Swedish sheep flocks was investigated. Finally, Section 2.7 concludes with a discussion.

2.3 Faecal Egg Count Reduction Test

The FECRT was suggested in the WAAVP guideline for estimating the reduction in FECs and its corresponding CI (Coles et al., 1992). Generally an analytical sensitivity of 50 is used, 15 is also possible if a greater sensitivity is desired. In order to reduce the counting variability, using groups of at least 10-15 animals was suggested. In addition, the mean pre-treatment FECs should be at least 150 epg, otherwise the FECRT can give unreliable results.

Suppose a group of n_T animals received anthelmintic treatment and a group of n_C animals serves as control. The percentage reduction in FECs can be calculated as

$$\text{Percentage reduction } = 100 \times \left(1 - \frac{\bar{x}_T}{\bar{x}_C}\right), \tag{2.3.1}$$

where $\bar{x}_T$ and $\bar{x}_C$ denote the mean counts of the treatment and the control group. As-

suming independence, the asymptotic variance of the log ratio is given by

$$\text{Var}\left(\log\frac{\bar{X}_T}{\bar{X}_C}\right) = \frac{1}{\bar{x}_T^2}\text{Var}(\bar{X}_T) + \frac{1}{\bar{x}_C^2}\text{Var}(\bar{X}_C). \tag{2.3.2}$$

The variance can be used to construct an approximate 95% CI of the log ratio using the 97.5% and the 2.5% quantile of a Student's t-distribution with $n_T + n_C - 2$ degrees of freedom. The CI for the log-ratio can be then transformed back to obtain the 95% CI for the estimated reduction. The WAAVP guideline (Coles et al., 1992) states that for sheep and goats, the resistance is present if (i) the percentage reduction in FECs is less than 95% and (ii) the corresponding lower 95% confidence limit is less than 90%. If only one of these two criterion is met, then a resistance is suspected. Different thresholds have been suggested for other livestock. Even if the anthelmintic treatment does not completely eradicate the infection, it can still be beneficial as a control program to maintain the productivity of livestock (Coles et al., 2006).

Over the recent years, modified versions of the FECRT have been proposed in literature. Wood et al. (1995) suggested to use the geometric mean in the FECRT instead of arithmetic mean. Davison and Hinkley (1997) suggested the 95% CI can also be calculated using nonparametric bootstrap. In the unpaired design, there is one group of animals that receives the treatment and another group is chosen to act as the control group. McKenna (1990) suggested that instead of taking samples from two groups of animals, the pre-treatment counts from the treatment group can be used as the baseline, hence eliminated the need of a distinct control group. We refer to this as the paired design. In this case, the FECRT becomes inappropriate since it does not take the paired structure into account.

2.4 Bayesian Hierarchical Models

There are two designs that can be used for detecting anthelmintic resistance in a livestock population. For each design, we propose a zero-inflated Bayesian hierarchical model to estimate the reduction in FECs.

2.4.1 The Unpaired Design

Suppose we have two groups of animals from the same population, a control group with size n_C and a treatment group with size n_T. A faecal sample from each animal is collected and counted using the McMaster technique with an analytical sensitivity f_i, where i is the index of each animal in the corresponding group. For notational simplicity, we assume every sample has the same analytical sensitivity, hence the index in f_i is dropped for the rest of the paper. The faecal sample is thoroughly mixed after dilution, hence we assume the eggs are homogeneously distributed within each sample. A proportion of the diluted sample $p = 1/f$ is then placed onto the McMaster chamber for counting. Denote the raw number of eggs in the diluted sample of the ith control animal as Y_i^{*C}, with $i = 1, 2, ..., n_C$. Given the true number of eggs per gram of faeces Y_i^C, the raw count Y_i^{*C} follows a binomial distribution with size Y_i^C and probability p. This captures both the dilution and the McMaster counting variability. Then the true epg Y_i^C follows a zero-inflated Poisson (ZIP) distribution with mean μ_i^C and zero-inflation parameter ϕ, it implies Y_i^C is 0 with probability ϕ, and follows the Poisson distribution with mean μ_i^C with

probability $(1-\phi)$. This captures the excess number of zero counts that could potentially comes from the unexposed animals. Finally the mean μ_i^C is gamma-distributed with shape κ and rate κ/μ. It has mean μ and variance μ^2/κ, the gamma distribution captures the overdispersion of the egg counts. This yields the following model for the control group animals,

$$\begin{aligned} Y_i^{*C}|Y_i^C &\sim \text{Bin}(Y_i^C, p), \\ Y_i^C|\mu_i^C, \phi &\sim \text{ZIP}(\mu_i^C, \phi), \\ \mu_i^C|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{2.4.1}$$

For the treatment group, the number of eggs in faecal samples is likely to decrease after some days receiving the treatment, but the treatment is unlikely to eliminate the infection in the animals completely, hence the zero-inflation component remains the same. In addition, we assume the reduction in FECs occurs at individual level, such that the parameters μ and κ also stay the same. We introduce a reduction factor $(1-\delta)$ where δ represents the proportion of eggs remaining. This yields the following model for the treatment group,

$$\begin{aligned} Y_i^{*T}|Y_i^T &\sim \text{Bin}(Y_i^T, p), \\ Y_i^T|\mu_i^T, \phi &\sim \text{ZIP}(\delta\mu_i^T, \phi), \\ \mu_i^T|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{2.4.2}$$

where the superscript T denotes the parameters for the treatment group. The priors for the flock parameters μ, κ and ϕ need to be specified in advance. If previous knowledge about the distribution of those parameters is available, they can be taken into account in the model as priors. Otherwise, diffuse priors should be used.

2.4.2 The Paired Design

In the paired design, there is only one group of animals of size n. A faecal sample from each animal is counted twice using the McMaster technique, once before the treatment and once some days after the treatment. The baseline counts of each animal is used as the corresponding control. The model for the paired design is

$$\begin{aligned} Y_i^{*C}|Y_i^C &\sim \text{Bin}(Y_i^C, p), \\ Y_i^C|\mu_i^C, \phi &\sim \text{ZIPois}(\mu_i^C, \phi), \\ Y_i^{*T}|Y_i^T &\sim \text{Bin}(Y_i^T, p), \\ Y_i^T|\mu_i^C, \phi &\sim \text{ZIPois}(\delta\mu_i^C, \phi), \\ \mu_i^C|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{2.4.3}$$

The only difference in the model comparing with the unpaired design is that, the pre-treatment epg Y_i^C and post-treatment epg Y_i^T are now based on the same Poisson mean μ_i^C to indicate that they belong to the same animal. The priors for the flock parameters should be specified in a similar way as for the unpaired design.

The hierarchical model (2.4.3) without zero-inflation in Y_i^C and Y_i^T was proposed in (Paul et al., 2014), however the authors used the posterior median as the point estimate for the reduction, and the 95% HPD credible interval as the interval estimate. The model was implemented in the `eggCounts` package (Paul, 2015) in R along with the hierarchical model for the unpaired design without zero-inflation. In addition, the authors used $(1-\bar{Y}_i^C/\bar{Y}_i^T)$

as the posterior samples for the reduction in the unpaired model rather than using $(1-\delta)$ directly. Typically, the posterior mode is used in conjunction with the HPD interval. In the simulation study, we show that using the posterior mode of the reduction parameter as the estimate gives a smaller bias compare to using the posterior median.

2.5 Simulation Study

In order to investigate the performance of the proposed Bayesian models, we conduct a simulation study to estimate the FECs reduction. We first simulate the FECs data under different scenarios, then use our proposed models and other methods to estimate the reduction. The bias and the coverage of the 95% CIs or credible intervals are compared across different methods.

2.5.1 Simulation Setup

FECs for both unpaired and paired designs are simulated. For each design, we consider 16 different scenarios that vary in terms of the baseline mean count μ (150 epg or 500 epg), the dispersion κ (1 or 2), the reduction $(1-\delta)$ (90% or 95%) and the zero-inflation ϕ (0 or 30%). Sample size is chosen to be 15 for all scenarios, and the analytical sensitivity is 50. For each scenario in each design, 1000 dataset are simulated. The pre-treatment FECs are simulated as follows: we firstly draw the mean epg μ_i^C from a gamma distribution with shape κ and rate κ/μ. Then the true number of eggs y_i^C are drawn from a ZIP distribution with mean μ_i^C and zero-inflation ϕ. Marginally, the true number of eggs follows a zero-inflated negative binomial distribution (Section 3.2. Finally, the observed counts are drawn from another Poisson distribution with mean y_i^C/f where f is the analytical sensitivity. The post-treatment FECs are simulated in a similar way but with different parameters. Note the process of simulating the data does not exactly match our proposed model, this encourages a fair comparison across the different methods.

We compare several different methods for estimating the mean FECs reduction and its confidence interval. For the unpaired design, we consider 1) the FECRT with the approximate CI (FECRT); 2) parametric bootstrap, assuming zero-inflated negative binomial distributions and using 1999 bootstrap samples (pBoot); 3) the hierarchical model in (2.4.1-2.4.2) without zero-inflation and using posterior median as the point estimate, as implemented in (Paul, 2015) (PoGa(median)) and 4) the same model but using posterior mode as the point estimate (PoGa(mode)); and finally 5) our proposed zero-inflated hierarchical model for the unpaired design (ZIPoGa).

The FECRT does not distinguish between paired and unpaired designs, hence it is applicable to both. The zero-inflated negative binomial regression does not perform well when the sample size is small, and it sometimes does not produce sensible results (Denwood et al., 2008). Hence for the paired design, in addition to 1) the FECRT, we consider 2) quasi-Poisson regression, excluding zero pre-treatment counts and using log pre-treatment counts as the offset term (qPois); 3) the proposed hierarchical model in (Paul et al., 2014) using posterior median as the point estimate (PoGa(median)) and 4) the same model but using posterior mode as the point estimate (PoGa(mode)); and finally 5) our proposed zero-inflated hierarchical model for the paired design (ZIPoGa).

The Bayesian models are implemented in the modelling language Stan (Carpenter, 2015), it uses an effective MCMC sampling technique and is available through the `rstan` package (Guo et al., 2015) in R (R Core Team, 2015). The prior for the reduction follows a Beta$(1, 1)$ distribution, which assigns uniform density between 0 and 1. For the parameters μ and κ, we use Gamma$(1, 0.001)$ and Gamma$(1, 0.7)$ prior respectively. For each Bayesian model, 12,000 MCMC samples are generated with 2,000 samples for burn-in without thinning. The posterior mode is used as the estimate for the reduction parameter in our proposed models, and the 95% HPD interval of the posterior samples was obtained. All the simulations are conducted in R version 3.2.1.

2.5.2 Simulation Results

Figure 2.2 and Figure 2.3 show the bias and the coverage probability of 95% CIs or 95% HPD interval for the FECs reduction, in the case of unpaired designs. The PoGa(median) model slightly underestimate the reduction in most cases, however it is improved by using the posterior mode as the point estimate as shown in PoGa(mode). All the other methods have small biases. Both the FECRT and the parametric bootstrap method have inaccurate coverage probabilities when the pre-treatment mean count is low. As expected, the FECRT has accurate coverage when the pre-treatment mean is high, since the asymptotic variance improves. The PoGa(median) model provides low coverage probability when the pre-treatment mean count is high, and it is improved by using $(1 - \delta)$ as the posterior samples for the reduction directly. In contrast, our proposed zero-inflation models offers good coverage while maintaining small bias. Note the Bayesian credible intervals do not have a long-run property like the CIs where 95 percent of the 95% CIs should cover the true parameter value (Spiegelhalter et al., 2004), but the coverage probability for the Bayesian methods can still be used as a rule of thumb to assess the models.

Figure 2.4 and Figure 2.5 show the bias and the coverage probability for the paired designs. The biases are small for all the methods except the PoGa(median) model. It is improved again by using the posterior mode as the estimate. In term of the coverage, the FECRT method tends to have wide confidence intervals since they do not take the paired structure into account, resulting almost 100% coverage when the pre-treatment mean is high. The Bayesian models provide slightly over-coverage in all the scenarios.

Overall, the zero-inflated Bayesian models are robust methods. They consistently provide small bias and accurate coverage in the simulated scenarios. In the following case study, we further illustrate the advantages of the zero-inflated hierarchical models.

2.6 Case Study: Anthelmintic Resistance in Swedish Sheep Flocks

In order to illustrate our proposed model, we re-analyze the data in a study where the prevalence of anthelmintic resistance in parasitic nematodes in Swedish sheep flocks was investigated (Höglund et al., 2009). The FECs data was collected and analyzed using both the FECRT and molecular testing methods. In the study, a total of 45 farms were randomly selected throughout Sweden, each with a minimum of 20 ewes. During the grazing season of 2006 and 2007, two flocks of approximately 15 lambs were selected from each farm, each flock was treated with either benzimidazole (BZ) or macrocyclic lactone. In this

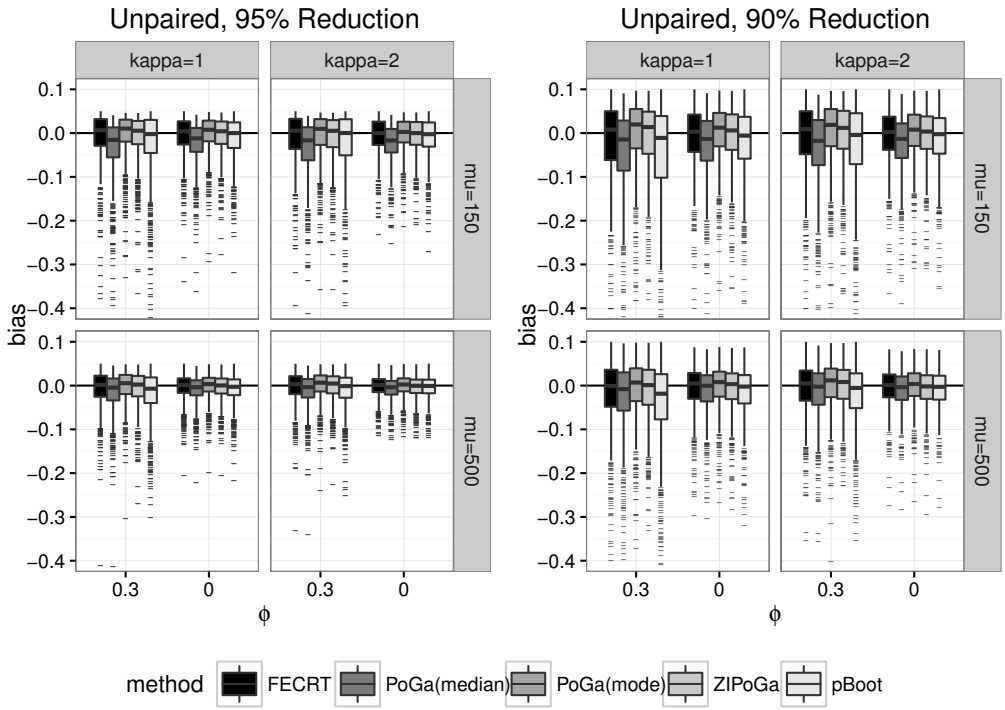


Figure 2.2: Boxplots of the estimated FECs reduction in the unpaired design, using 1) FECRT with approximated CI; 2) the hierarchical model without zero-inflation (Paul, 2015) and using the posterior median of $(1-\bar{Y}_i^C/\bar{Y}_i^T)$ as the point estimate; 3) the hierarchical model without zero-inflation (Paul, 2015) and using the posterior mode of $(1-\delta)$ as the point estimate; 4) our proposed zero-inflated hierarchical model for the unpaired design; and 5) parametric bootstrap. The horizontal line indicates zero bias.

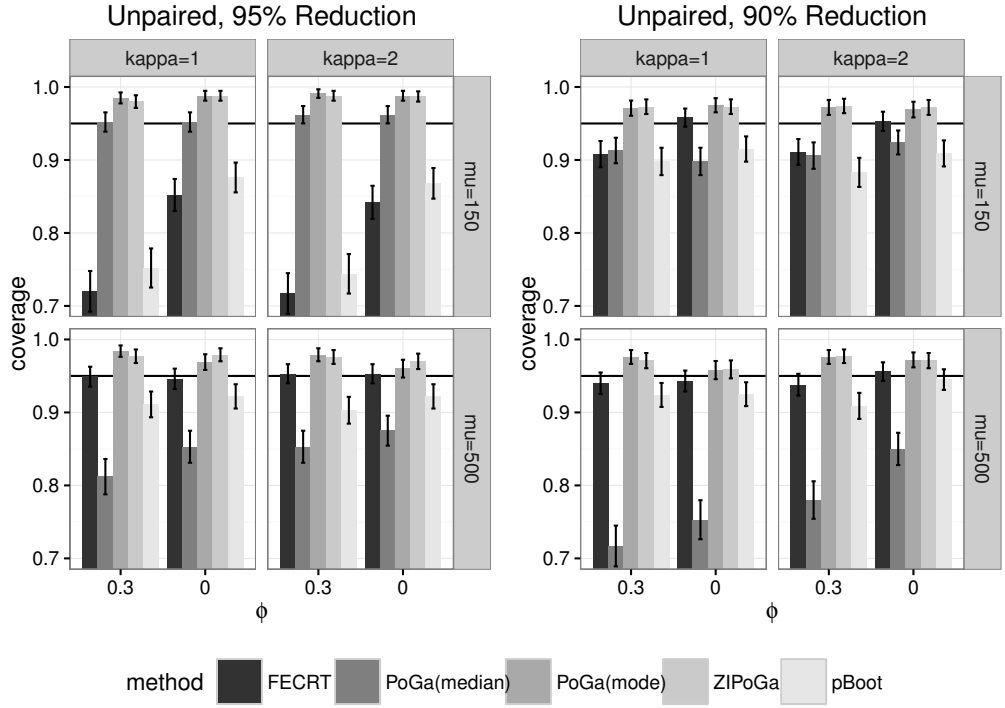


Figure 2.3: Barplots of the coverage probability of the 95% CIs, or HPD credible intervals for the FECs reduction in the unpaired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Figure 2.2.

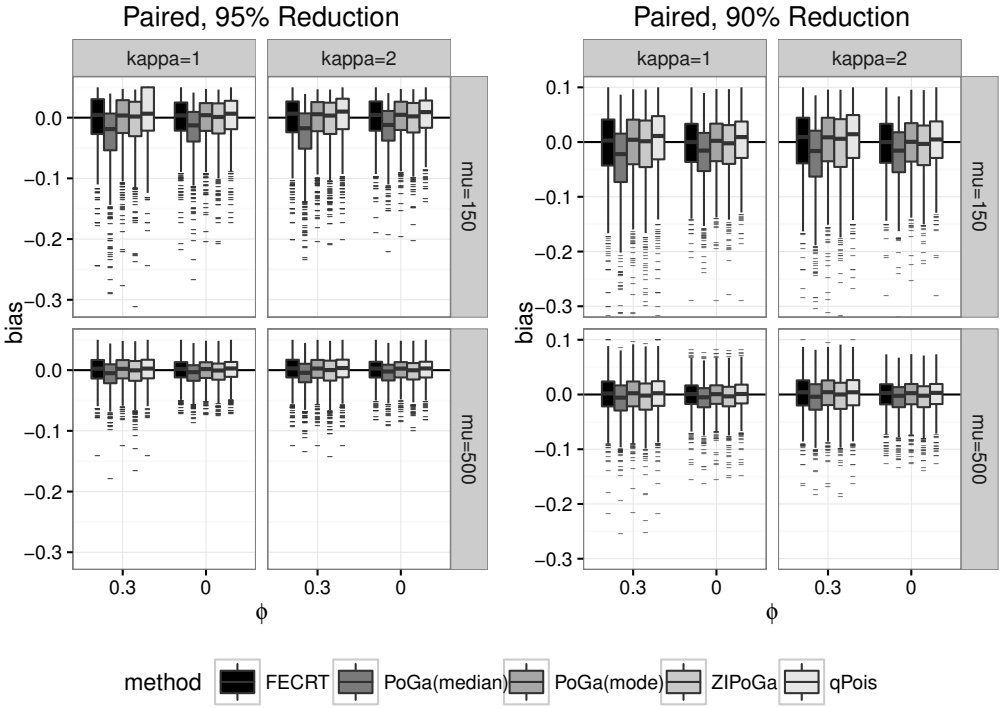


Figure 2.4: Boxplots of the estimated FECs reduction in the unpaired design, using 1) FECRT with approximated CI; 2) the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior median as the point estimate; 3) the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior mode as the point estimate; 4) our proposed zero-inflated hierarchical model for the paired design; and 5) quasi-Poisson regression. The horizontal line indicates zero bias.

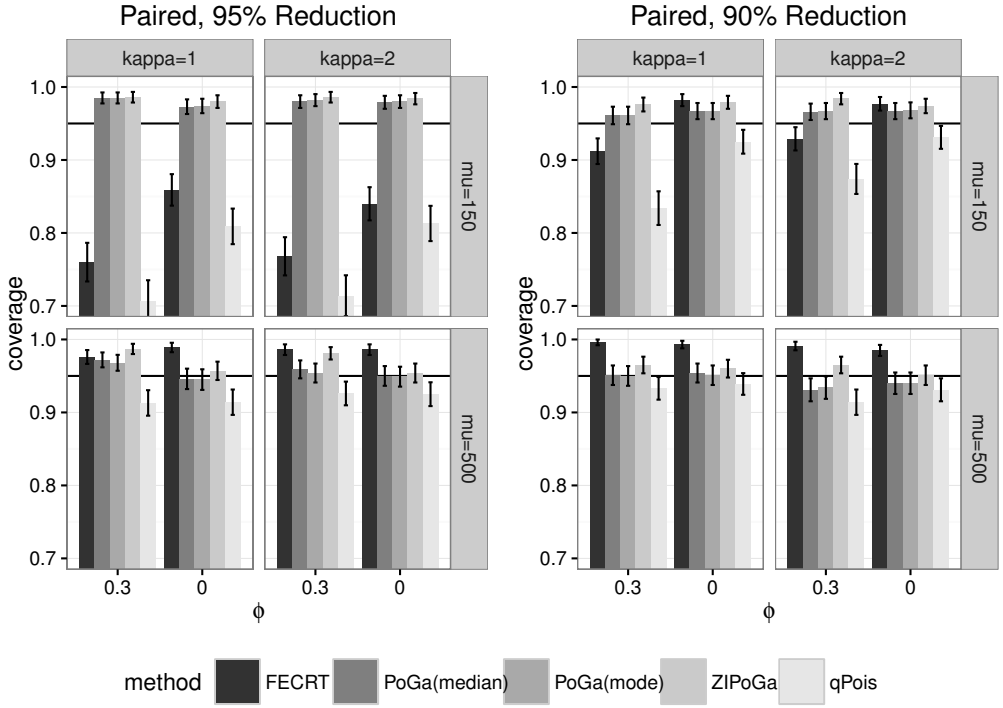


Figure 2.5: Barplots of the coverage probability of the 95% CIs, or HPD credible intervals in the case of Bayesian models, for the FECs reduction in the paired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Figure 2.4.

paper, we only consider the efficacy of BZ, which was received by 45 out of all 90 flocks selected. However the model is applicable for other treatments as well as other livestock. Each lamb was sampled before treatment using the modified McMaster technique with an analytical sensitivity of 50. 39 out of 45 flocks with mean of at least 50 epg was re-sampled using the same setting 7-10 days after treatment, with flock sizes varying between 10 to 17 animals. In addition to the McMaster counting technique, the BZ-resistance of parasites was tested using Pyrosequencing assay. Larval cultures indicated that Teladorsagia and Trichostrongylid nematode infection were predominant.

There are 39 flocks consisting of 575 animals in total, all of them were treated with BZ. The post-treatment FECs are missing in 28 animals, hence they are excluded from the analysis. In addition, one animal had a pre-treatment epg of 30, which is not possible with a correction factor of 50. In this case, the author clarified that 3 eggs were observed outside the grid area on the McMaster slide, hence a correction factor of 10 was applied. However according to WAAVP guideline, eggs outside the grid should not be counted, hence this particular observation was set to zero. Using FECRT, we first calculate the reduction in FECs and its approximate 95% CI. The decision rule for sheep and goats suggested in the WAAVP guidelines is used for deciding anthelmintic resistance. In 35 flocks, all of the post-treatment counts are zero which resulted 100% reduction in each flock. The CI for those flocks cannot be computed using the FECRT. Out of the remaining 4 flocks, 2 flocks (flock 33 and 39) are anthelmintic resistant according to the FECRT. The results based on the molecular testing suggested 5 out of 39 flocks (flock 24, 33, 36, 37 and 39) have anthelmintic resistance present using the codon 200 TAC allele frequency of $\geq 95\%$ as the indicator. In the end, the authors concluded that the prevalence of anthelmintic resistance in the Swedish sheep population is relatively low, however it is more widespread than the FECRT indicated. The paper pointed out the urgent need to develop alternative diagnostic procedure. The quasi-Poisson regression gave similar results, failing to detect the remaining resistance.

In the following, we re-analyze the FECs data from the Swedish sheep study using our proposed model. The worm burden differs depending on the animals and the type of parasites eggs that is being counted, hence the choice of hyperparameters for the prior should be based on similar studies. According to another study of the distribution of trichostrongylid eggs in the sheep flocks (Morgan et al., 2005), the mean pre-treatment FECs ranged from 43 to 1915, and the estimated dispersion parameter based on negative binomial regressions ranged from 0.18 (95% CI: 0.10 to 0.32) to 2.3 (95% CI: 0.2 to 4.2). Hence we assign a weakly informative prior Gamma(1, 0.001) to μ, where 90% of the probability mass lies between 59 and 2996, and assign a Gamma$(1, 0.7)$ prior for κ, where 90% of the probability mass lies between 0.1 and 4.3 with a prior median of 1. We assume the overall level of infection does not increase after treatment is applied, hence the reduction should always be between 0 to 100%. A non-informative prior Beta$(1, 1)$ is assigned to the parameter δ, such that all the values between 0 and 1 are equally likely a priori. Finally for the zero-inflation parameter ϕ, we assign a non-informative Beta$(1, 1)$ prior.

We apply the zero-inflated Bayesian model for the paired design separately to each flock. In order to diagnose the potential non-convergence, 4 MCMC chains were requested. Each has 12,000 MCMC samples, 2,000 samples for burn-in and without thinning. There was no evidence of non-convergence with potential scale reduction factors (Brooks and Gelman, 1998) approximately equal to 1. The sensitivity analysis showed similar results with wide

Flock	FECRT (CI)	qPois	PoGa(mode)	pZIPoGa
24	99.0 (96.3, 99.8)	99.0 (97.2, 99.7)	99.0 (98.5, 99.4)	97.8 (95.8, 98.9)
33	**82.2 (65.4, 90.8)**	**82.2 (68.6,90.0)**	**81.3 (77.4, 85.9)**	**76.8 (70.6,81.8)**
36	97.5 (90.6, 99.4)	97.5 (93.2,99.1)	97.6 (93.1, 99.2)	97.4 (93.1, 99.2)
37	100.0 (-, -)	100.0 (100.0, 100.0)	*99.3 (89.5, 100.0)*	*98.8 (49.3, 100.0)*
39	**92.3 (62.9, 98.4)**	**93.9 (90.1, 96.3)**	**92.6 (89.0, 94.8)**	**93.1 (89.7, 95.6)**

Table 2.1: Analysis results for the five BZ treated flocks which the molecular testing indicated anthelmintic resistance are present. Results are shown for the estimated percentage reduction in FECs using FECRT and its approximate 95% CI. Following is the results from the quasi-Poisson regressions with 95% CI. The last two columns are the Bayesian hierarchical models PoGa(mode) and ZIPoGa. The posterior mode and 95% HPD intervals are shown for each model. The text is in **bold** if a resistance is detected, and is in *italic* if a resistance is suspected.

uniform priors on the mean and dispersion, here we only present the main results. Table 2.1 shows the results for the five flocks which the molecular data indicated anthelmintic resistance. The approximate CI cannot be computed for flock 37 using the FECRT, since all the post-treatment FECs are zero. Because the standard FECRT method does not take the paired structure into account, the approximate CI is wider in general compares to the quasi-Poisson regression and the Bayesian models. The Bayesian models are able to obtain an interval estimate even when the reduction is 100%. The posterior mode estimate for the Bayesian model without zero-inflation is similar to the FECRT, however the zero-inflated Bayesian model gave slightly different estimates. In particular, the posterior mode for the reduction in flock 33 is 76.8% using our proposed model, compare to 82.2% and 81.3% in the FECRT and PoGa(mode). Indeed, the mean reduction calculated using Eq (2.3.1) is 82.2%, however this completely ignores the paired structure. The actual mean pairwise reduction for flock 33 is 73.1%, hence our proposed ZIPoGa model provide a more sensible result in this case. For flock 37, the Bayesian models classify it as suspected resistance due to its lower confidence limit. Since no parasite eggs were detected in 7 out of 13 sheep before treatment, the uncertainty in the treatment efficacy is high. Hence the interval estimate is much wider, which is only captured by the zero-inflation model. The other classification results stay the same.

Figure 2.6 shows the estimated reductions and its 95% HPD intervals for all 39 flocks considered in the case study. There are several flocks that are flagged as suspected resistance even though there were no eggs present in the post-treatment FECs. For example, flock 35 has 15 sheep, all of which had zero post-treatment FECs. However, 10 out of 15 sheep had zero pre-treatment counts, those could be the unexposed individuals that should not contribute to the estimation of treatment efficacy. This is captured by the zero-inflated model, hence the HPD credible interval for this flock is wide.

Another advantage of the Bayesian approach is that it does not only provide the reduction estimate and the credible interval, but also it offers density distributions of the model parameters. (Denwood et al., 2010) pointed out that Bayesian methods allow for probabilistic classification on the efficacy, in terms of the probability that a true reduction is below a given threshold. According to the WAAVP guidelines, there are three possible decision outcomes on resistance status, namely "yes", "suspected" and "no". Such trichotomy outcome should be interpreted with caution, especially at the decision boundaries. We illustrate the probabilistic view using flock number 37 and 39. Figure 2.7 shows the posterior marginal density of the reduction parameter $(1 - \delta)$ from the proposed model. Coles et al. (2006) stated that a reduction greater than 95% is considered as beneficial, hence we use this as the threshold. The shaded area in each case corresponds to the probability that

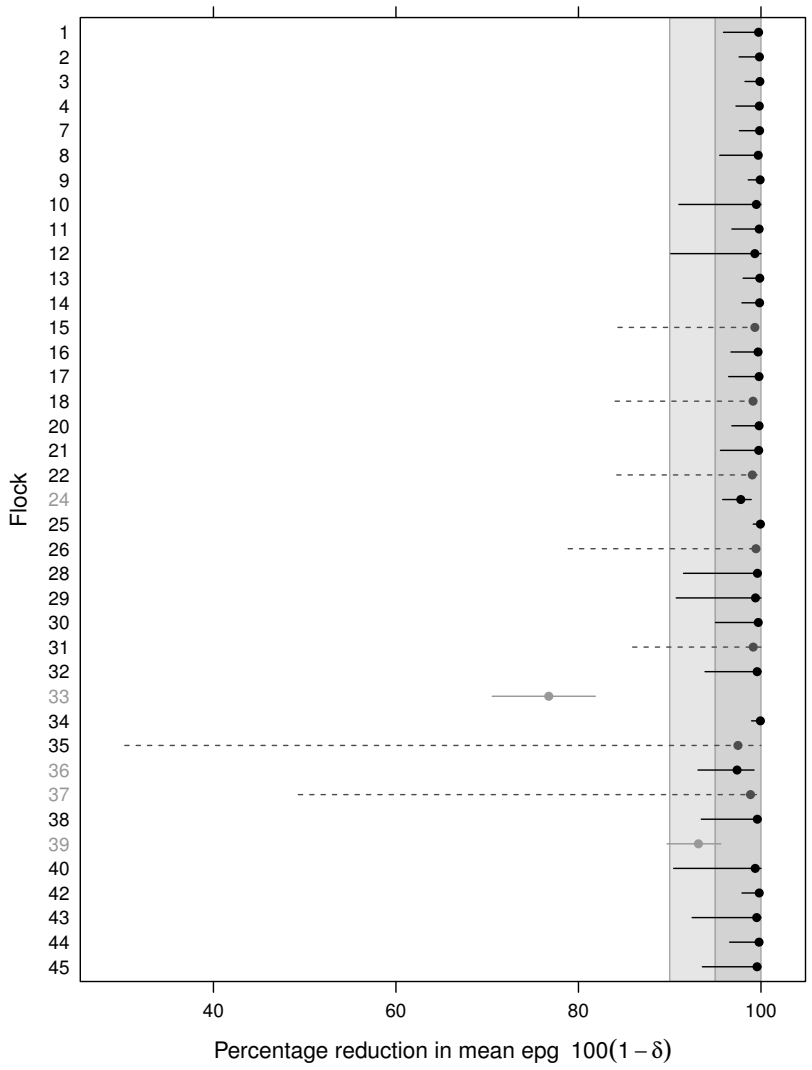


Figure 2.6: Estimated reduction in mean FECs and its 95% HPD interval for the 39 flocks that were sampled both before and after treated with BZ. Using the WAAVP guideline for the decision of anthelmintic resistance, the intervals in solid black lines belong to flocks with no anthelmintic resistance, intervals in dashed lines belong to flocks with suspected resistance and intervals in solid gray lines belong to flocks with resistance. The flock numbers that were flagged as resistant using molecular data are colored in grey.

the reduction in mean FECs is less than 95%, i.e. the probability that anthelmintic resistance is present using a 95% reduction as the threshold. Based on the posterior marginal distribution, the probability that the resistance is present in flock 37 is 0.42, indicating moderate evidence for resistance. For flock 39, the probability is 0.94 which suggests a very strong evidence that the resistance is present.

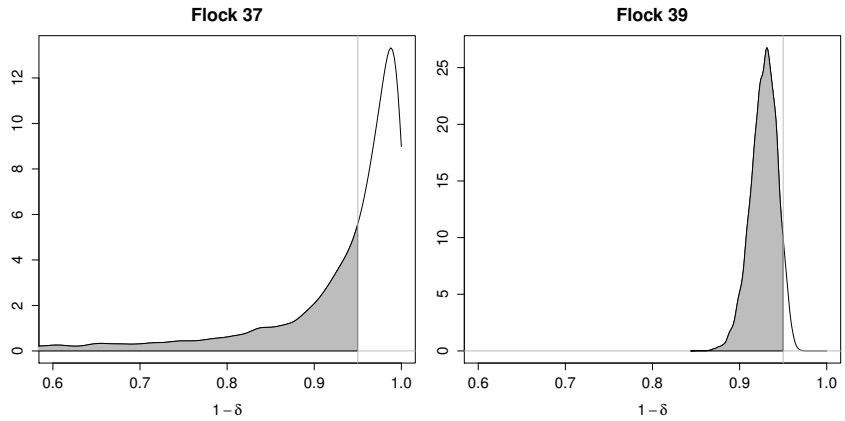


Figure 2.7: The marginal posterior density for the reduction (1-δ) for flock 37 and 39. The shaded area represents the density mass for reduction less than 95%.

2.7 Discussion

The anthelmintic resistance of gastrointestinal nematodes has posed a serious problem to the ruminant production in recent years. The infection level is often quantified by the FECs in animal faeces. The standard method FECRT estimates the reduction in FECs and its approximate CI, however it only computes the mean counts and ignores the additional variations. As a result, the FECRT only produces reliable estimates for high infection levels. Moreover, as illustrated in the case study, the FECRT provides an estimate of 100% reduction without a CI when all the post-treatment FECs are zero. In this paper we propose zero-inflated Bayesian hierarchical models for estimating the reduction in FECs. The models capture the additional sources of variabilities in the data, and allow for extra zero counts that is frequently observed in practice due to unexposed animals. The simulation results suggest that the zero-inflated Bayesian hierarchical models are robust methods to estimate the reduction, in both unpaired and paired designs. They consistently provide small bias and good coverage in all the simulated scenarios. The case study further illustrated the advantages of our proposed model, which can accurately model the paired structure and provide an interval estimate where the conventional method cannot.

Another advantage of the Bayesian hierarchical models is its flexibility in model formulation. In this paper we have assumed the reduction in FECs is the same for every animal, as one can expect the efficacy of anthelmintic treatment across animals are similar within a resistant community. However each animal can experience different efficacy due to different metabolism or drug availability (Cabaret and Berrag, 2004), one can adjust the model to introduce animal-specific reductions. Sufficient data are required to ensure the convergence of the model. In the case study, if researchers are interested in assessing the anthelmintic resistance in the Swedish sheep population in general, a hierarchical meta-

analysis model over all the flocks can be formulated. The corresponding model parameters from each flock would follow the same distributions, for example, the parameter μ from each flock together follows a normal distribution with some population mean. This can be particular useful if one wish to consider some national treatment schemes applied to the entire sheep population.

With the proposed models in mind, one can also design more efficient sampling process in order to obtain the estimated FECs reduction with sufficient statistical power. The sample size and the analytical sensitivity are the two important factors involved in a study design. The CIs are expected to be narrower for larger sample size and higher analytical sensitivity. Sample size calculation problem and the choice of analytical sensitivity can be further investigated for the zero-inflated Bayesian models.

Chapter 3

Additional Proofs

3.1 Relation between Faecal Egg Count Reduction Test and Poisson Regression

Although the conventional FECRT is straight-forward to calculate, the estimated FECs reduction is mathematically equivalent to the estimate from Poisson regression.

If we model the post-treatment FECs with a Poisson regression, with log link function, intercept β_0 without any covariates and logarithm of the pre-treatment FECs $\log(y_i^C)$ as the offset term, we obtain

$$\log(\mu_i) = \beta_0 + \log(y_i^C), \tag{3.1.1}$$

where μ_i is the Poisson mean of the ith observation. The maximum likelihood estimator (MLE) of β_0 in the Poisson regression is the same as the MLE of a quasi-Poisson regression. In order to apply the offset term, any zero pre-treatment FECs and their corresponding post-treatment FECs should be excluded. If both pre- and post-treatment FECs are zero, then there will be no influence on the total number of counts. However, this is not the case when pre-treatment FECs are zero while the corresponding post-treatment FECs are not. Therefore, this equivalence only holds true if the corresponding post-treatment FECs are also zero.

In general, the Poisson regression assumes $Y_i|x_i \sim \text{Pois}(\mu_i)$, and has the relationship $\log(\mu_i) = x_i\beta$. The likelihood contribution of the observations are

$$L(\beta) = \prod_{i=1}^{n} \frac{\mu_i^{y_i}}{y_i} e^{-\mu_i}. \tag{3.1.2}$$

Substituting $\mu_i = \exp(\beta_0 + \log(y_i^C))$ and use y_i^T as the response to replace y_i, the log-likelihood becomes

$$l(\beta_0) = \sum_{i=1}^{n} y_i^T \left(\beta_0 + \log(y_i^C)\right) - \log(y_i^T) - \exp\left(\beta_0 + \log(y_i^C)\right). \tag{3.1.3}$$

The maximum likelihood estimator for β_0 can be obtained by setting the first derivative

of the log-likelihood to be zero and solve for β_0 as

$$\begin{aligned} l'(\beta_0) &= \sum_{i=1}^{n} y_i^T - \exp(\beta_0 + \log(y_i^C)) = 0, \\ \sum_{i=1}^{n} y_i^T &= \sum_{i=1}^{n} \exp(\hat{\beta}_0) y_i^C, \\ \exp(\hat{\beta}_0) &= \frac{\sum_{i=1}^{n} y_i^T}{\sum_{i=1}^{n} y_i^C} = \frac{\bar{y}_i^T}{\bar{y}_i^C}. \end{aligned} \tag{3.1.4}$$

Therefore, $1 - \exp(\hat{\beta}_0)$ is equivalent to the percentage reduction in Eq (2.3.1).

3.2 Equivalence of Zero-Inflated Negative Binomial Formulations

In this section, we show the mixture of a zero-inflated Poisson distribution and a gamma distribution is equivalent to a ZINB distribution. We use the pre-treatment FECs Y_i^C as an example. For notational simplicity, we drop the superscript C. We denote $\mathbb{1}$ to be an indicator function. According to model (2.4.1), we have

$$\begin{aligned} f(Y_i) &= \int_0^\infty f_{\text{ZIPoisson}}(y_i; \mu_i, \phi) \times f_{\text{Gamma}}(\mu_i; \kappa, \kappa/\mu)\, \mathrm{d}\mu_i \\ &= \int_0^\infty \mathbb{1}\{y_i \neq 0\}(1-\phi)\frac{{\mu_i}^{y_i} e^{-\mu_i}}{y_i!} \times \frac{(\kappa/\mu)^\kappa e^{-\kappa/\mu \times \mu_i} {\mu_i}^{\kappa-1}}{\Gamma(\kappa)} \\ &\quad + \mathbb{1}\{y_i = 0\}\left(\phi + (1-\phi)e^{-\mu_i}\right) \frac{(\kappa/\mu)^\kappa e^{-\kappa/\mu \times \mu_i} {\mu_i}^{\kappa-1}}{\Gamma(\kappa)}\, \mathrm{d}\mu_i \\ &= \mathbb{1}\{y_i \neq 0\} \times (1-\phi)\frac{(\kappa/\mu)^\kappa}{y_i!\Gamma(\kappa)} \int_0^\infty {\mu_i}^{y_i+\kappa-1} e^{-\kappa/\mu \times \mu_i - \mu_i}\, \mathrm{d}\mu_i \\ &\quad + \mathbb{1}\{y_i = 0\} \times \left(\phi + (1-\phi)\frac{(\kappa/\mu)^\kappa}{\Gamma(\kappa)} \int_0^\infty {\mu_i}^{\kappa-1} e^{-\kappa/\mu \times \mu_i - \mu_i}\, \mathrm{d}\mu_i\right) \\ &= \mathbb{1}\{y_i \neq 0\} \times (1-\phi)\frac{(\kappa/\mu)^\kappa}{y_i!\Gamma(\kappa)}\Gamma(\kappa + y_i)\left(\frac{\mu}{\kappa+\mu}\right)^{\kappa+y_i} \\ &\quad + \mathbb{1}\{y_i = 0\} \times \left(\phi + (1-\phi)\frac{(\kappa/\mu)^\kappa}{\Gamma(\kappa)}\Gamma(\kappa)\left(\frac{\mu}{\kappa+\mu}\right)^{\kappa}\right) \\ &= \mathbb{1}\{y_i \neq 0\}(1-\phi) f_{\text{NB}}(y_i; \mu, \kappa) + \mathbb{1}\{y_i = 0\} \times (\phi + (1-\phi)(1+\mu/\kappa)^{-\kappa}) \\ &= (1-\phi) f_{\text{NB}}(y_i; \mu, \kappa) + \mathbb{1}\{y_i = 0\}\phi, \end{aligned}$$

Hence if the true epg Y_i is 0, then it can come from both the negative binomial distribution with probability $(1-\phi)$ and zero mixture with probability ϕ being in the mixture. If it is not 0, then it can only come from the negative binomial distribution. This is equivalent to a ZINB distribution.

Chapter 4

Summary

This thesis has studied the statistical modelling of zero-inflated faecal egg counts in order to assess the anthelmintic efficacy. The Bayesian hierarchical models have been proposed for estimating the reduction in faecal egg counts, for both unpaired and paired designs. In a simulation study, the bias and coverage probability of 95% HPD credible intervals have been compared with the existing methods.

This thesis began with the epidemiology of gastrointestinal nematodes in livestock, and the current statistical models for analyzing faecal egg counts data. The previous work by Paul et al. (2014) was described and the contributions of this thesis were listed. It followed by the working paper *Zero-Inflated Hierarchal Models for Faecal Egg Counts to Assess Anthelmintic Efficacy*, which focused on the newly proposed zero-inflation Bayesian models. The results from a simulation study showed, our proposed hierarchical models provide robust estimate of the reduction in faecal egg counts, they behave well in terms of the bias and the coverage probability. A case study on the anthelmintic efficacy of Swedish sheep flocks was then used to further demonstrate the advantages of the proposed model for the paired design. It captures the paired structure and provides sensible interval estimates where the conventional method cannot. Afterward, two proofs were shown. The first proof showed the estimate of FECs reduction using FECRT and the MLE from a Poisson regression are equivalent. The second proof showed the equivalence of the proposed model formulation with a ZINB distribution.

This thesis offered a probabilistic view on the reduction of faecal egg counts. Like the conventional FECRT, it is able to provide an estimate with a confidence interval for the reduction. However, it also provides a density distribution for the parameters of interest. The current decision rule for declaring anthelmintic resistance is based only on the mean estimate and the lower confidence limit of the FEC reduction. Alternative decision rules can be developed based on the information provided by the posterior distributions, preferably in consultation with researchers in parasitology.

The model can be extended in several ways. Firstly, we have assumed the reduction in FECs is the same for every animal, as one can expect the efficacy of anthelmintic treatment across animals are similar within a resistant community. However each animal can experience different efficacy due to different metabolism or drug availability (Cabaret and Berrag, 2004), one can adjust the model to introduce animal-specific reductions. Sufficient data are required to ensure the convergence of the model. In the case study, if researchers are interested in assessing the anthelmintic resistance in the Swedish sheep population in

general, a hierarchical meta-analysis model over all the flocks can be formulated. The corresponding model parameters from each flock would follow the same distributions, for example, the parameter μ from each flock together follows a normal distribution with some population mean. This can be particular useful if one wish to consider some national treatment schemes applied to the entire sheep population.

Lastly, there are other models that can be compared with. Levecke et al. (2015) used a Poisson distribution for the raw counts instead of the binomial distribution used in this thesis. This will involve a discrete sampling process which is not yet available in Stan. The differences between the models can be investigated in the future using other tools, with combination of the zero-inflation components which was not included in their model. An alternative R package `bayescount` by Denwood (2015) also implements Bayesian hierarchical models to analyze FECs data. It uses JAGS for the MCMC sampling, which also saves a lot of programming effort but it has a slower convergence speed compared with Stan. The exact model specification is also unclear, only the JAGS code is available in a PhD thesis (Denwood, 2010) without any publications on the models themselves. One of the model without zero-inflation (Denwood et al., 2010) was later published and compare with existing methods, it showed the Bayesian models have superior performance compare to both the FECRT and bootstrapping method. The zero-inflation models in the `bayescount` package can also be investigated further.

Bibliography

Alexander Neal (2012). Analysis of Parasite and Other Skewed Counts. *Tropical medicine & international health 17*(6), 684–693.

Brooks, S. P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics 7*(4), 434–455.

Cabaret, J. and B. Berrag (2004). Faecal egg count reduction test for assessing anthelmintic efficacy: average versus individually based estimations. *Veterinary Parasitology 121*(1 - 2), 105 – 113.

Carpenter, B. (2015). Stan: A probabilistic programming language. *Journal of Statistical Software*. In press.

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson (2016). *shiny: Web Application Framework for R*. R package version 0.13.1.

Coles, G., C. Bauer, F. Borgsteede, S. Geerts, T. Klei, M. Taylor, and P. Waller (1992). World association for the advancement of veterinary parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Veterinary Parasitology 44*(1 - 2), 35 – 44.

Coles, G., F. Jackson, W. Pomroy, R. Prichard, G. von Samson-Himmelstjerna, A. Silvestre, M. Taylor, and J. Vercruysse (2006). The detection of anthelmintic resistance in nematodes of veterinary importance. *Veterinary Parasitology 136*(3-4), 167 – 185.

Davison, A. and D. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Denwood, M. (2015). *bayescount: Power Calculations and Bayesian Analysis of Count Distributions and FECRT Data using MCMC*. R package version 0.9.99-5.

Denwood, M., S. Reid, S. Love, M. Nielsen, L. Matthews, I. McKendrick, and G. Innocent (2010). Comparison of three alternative methods for analysis of equine faecal egg count reduction test data. *Preventive Veterinary Medicine 93*(4), 316 – 323.

Denwood, M. J. (2010). *A quantitative approach to improving the analysis of faecal worm egg count data*. Ph. D. thesis, University of Glasgow. url:http://theses.gla.ac.uk/1837/.

Denwood, M. J., M. J. Stear, L. Matthews, S. W. J. Reid, N. Toft, and G. T. Innocent (2008). The distribution of the pathogenic nematode nematodirus battus in lambs is zero-inflated. *Parasitology 135*, 1225–1235.

Dobson, R., B. Hosking, C. Jacobson, J. Cotter, R. Besier, P. Stein, and S. Reid (2012). Preserving new anthelmintics: A simple method for estimating faecal egg count reduc-

tion test (FECRT) confidence limits when efficacy and/or nematode aggregation is high. *Veterinary Parasitology 186*(1-2), 79 – 92.

Donato, T., von Samson-Himmelstjerna Georg, Demeler Janina, Milillo Piermarino, Schürmann Sandra, Barnes Helen, Otranto Domenico, Perrucci Stefania, di Regalbono Antonio Frangipane, Beraldo Paola, Boeckh Albert, and Cobb Rami (2009). Anthelmintic resistance in cyathostomin populations from horse yards in Italy, United Kingdom and Germany. *Parasites & Vectors 2*(Suppl 2).

FAO (2015). Food And Agriculture Organization of the United Nations Statistics Division. http://faostat3.fao.org/browse/Q/QL/E. Accessed: 2016-01-27.

George Casella, E. I. G. (1992). Explaining the Gibbs sampler. *The American Statistician 46*(3), 167–174.

Grenfell, B. T., K. Wilson, V. S. Isham, H. E. G. Boyd, and K. Dietz (1995). Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology 111*, S135–S151.

Guo, J., J. Gabry, and B. Goodrich (2015). *rstan: R Interface to Stan*. R package version 2.8.2.

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

Höglund, J., K. Gustafsson, B.-L. Ljungström, A. Engström, A. Donnan, and P. Skuce (2009). Anthelmintic resistance in Swedish sheep flocks based on a comparison of the results from the faecal egg count reduction test and resistant allele frequencies of the β-tubulin gene. *Veterinary Parasitology 161*(1-2), 60 – 68.

Homan, M. D. and A. Gelman (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res. 15*(1), 1593–1623.

Houtert, M. F. V. and A. R. Sykes (1996). Implications of nutrition for the ability of ruminants to withstand gastrointestinal nematode infections. *International Journal for Parasitology 26*(11), 1151 – 1167.

Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends in Parasitology 20*(10), 477 – 481.

Levecke, B., R. M. Anderson, D. Berkvens, J. Charlier, B. Devleesschauwer, N. Speybroeck, J. Vercruysse, and S. V. Aelst (2015). Chapter five - Mathematical inference on helminth egg counts in stool and its applications in mass drug administration programmes to control soil-transmitted helminthiasis in public health. In *Mathematical Models for Neglected Tropical Diseases: Essential Tools for Control and Elimination, Part A*, Volume 87 of *Advances in Parasitology*, pp. 193 – 247. Academic Press.

Levecke, B., R. Dobson, N. Speybroeck, J. Vercruysse, and J. Charlier (2012). Novel insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance. *Veterinary Parasitology 188*(3–4), 391–396.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhaltery (2000). Winbugs - a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 325–337.

McKenna, P. (1990). The detection of anthelmintic resistance by the faecal egg count reduction test: An examination of some of the factors affecting performance and interpretation. *New Zealand Veterinary Journal 38*(4), 142–147.

Miller, C., T. Waghorn, D. Leathwick, and M. Gilmour (2006). How repeatable is a faecal egg count reduction test? *New Zealand Veterinary Journal 54*(6), 323–328.

Morgan, E., L. Cavill, G. Curry, R. Wood, and E. Mitchell (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep. *Veterinary Parasitology 131*(1-2), 79 – 87.

Mortensen, L. L., L. H. Williamson, T. H. Terrill, R. A. Kircher, M. Larsen, and R. M. Kaplan (2003). Evaluation of prevalence and clinical implications of anthelmintic resistance in gastrointestinal nematodes in goats. *Journal of the American Veterinary Medical Association 223*(4), 495–500.

Nødtvedt Ane, Dohoo Ian, Sanchez Javier, Conboy Gary, DesCôteaux Luc, Keefe Greg, Leslie Ken, and Campbell John (2002). The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows. *Canadian Journal of Veterinary Research 66*(4), 249–257.

Paul, M. (2015). *eggCounts: Hierarchical Modelling of Faecal Egg Counts*. R package version 0.4-1.

Paul, M., P. R. Torgerson, J. Höglund, and R. Furrer (2014). Hierarchical modelling of faecal egg counts to assess anthelmintic efficacy. *ArXiv e-prints*.

Perry, B. and T. Randolph (1999). Improving the assessment of the economic impact of parasitic diseases and of their control in production animals. *Veterinary Parasitology 84*(3-4), 145–168.

Pfukenyi, D., A. Willingham, S. Mukaratirwa, and J. Monrad (2007). Epidemiological studies of parasitic gastrointestinal nematodes, cestodes and coccidia infections in cattle in the highveld and lowveld communal grazing areas of Zimbabwe. *Onderstepoort J Vet Res 74*(2).

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. Available at https://www.r-project.org/conferences/DSC-2003/Drafts/Plummer.pdf.

Pullan, R. L., N. B. Kabatereine, R. J. Quinnell, and S. Brooker (2010). Spatial and genetic epidemiology of hookworm in a rural community in Uganda. *PLoS Neglected Tropical Diseases 4*(6), 1–10.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Slocombe, J. O. D., J. F. Coté, and R. V. de Gannes (2008, January). The persistence of benzimidazole-resistant cyathostomes on horse farms in Ontario over 10 years and the effectiveness of ivermectin and moxidectin against these resistant strains. *The Canadian Veterinary Journal 49*(1), 56–60.

Spiegelhalter, D. J., K. R. Abrams, and J. P. Myles (2004). *An Overview of the Bayesian Approach*, pp. 64–67. John Wiley & Sons, Ltd.

Tariq, K. A. (2014). A review of the epidemiology and control of gastrointestinal nematode infections of small ruminants. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences 85*(2), 693–703.

Torgerson, P., M. Schnyder, and H. Hertzberg (2005). Detection of anthelmintic resistance: a comparison of mathematical techniques. *Veterinary Parasitology 128*(3-4), 291 – 298.

Torgerson, P. R., M. Paul, and R. Furrer (2014). Evaluating faecal egg count reduction using a specifically designed package 'eggcounts' in R and a user friendly web interface. *International Journal for Parasitology 44*(5), 299 – 303.

Torgerson, P. R., M. Paul, and F. I. Lewis (2012). The contribution of simple random sampling to observed variations in faecal egg counts. *Veterinary Parasitology 188*(3-4), 397 – 401.

Waller, P. J. (1997). Anthelmintic resistance. *Veterinary Parasitology 72*(3 - 4), 391 – 412. Fourth Ostertagia Workshop: Nematode Parasites of Importance to Ruminant Livestock.

Waruiru, R., S. Thamsborg, P. Nansen, N. Kyvsgaard, H. Bogh, W. Munyua, and J. Gathuma (2001). The epidemiology of gastrointestinal nematodes of dairy cattle in central Kenya. *Tropical Animal Health and Production 33*(3), 173–187.

Wood, I., N. Amaral, K. Bairden, J. Duncan, T. Kassai, J. M. Jr., J. Pankavich, R. Reinecke, O. Slocombe, S. Taylor, and J. Vercruysse (1995). World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) second edition of guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine, ovine, caprine). *Veterinary Parasitology 58*(3), 181–213.

Zanzani, S. A., A. L. Gazzonis, A. Di Cerbo, M. Varady, and M. T. Manfredi (2014). Gastrointestinal nematodes of dairy goats, anthelmintic resistance and practices of parasite control in Northern Italy. *BMC Veterinary Research 10*(1), 1–10.

Appendix A

Stan Codes

This appendix includes the Bayesian model specifications in the Stan language. The codes can be used as character strings in the `mode_code` argument of `stan(...)` function in the `rstan` package. Alternatively, the models can be directly used using `fecr_stan(...)` function in the `eggCounts` package.

Zero-Inflated Bayesian Model for the Unpaired Design

```
data {
  int Ja; // number of animals
  int Jb;
  int ystararaw[Ja]; // after treatment McMaster count
  int ystarbraw[Jb]; // before treatment McMaster count
  int fpost[Ja];
  int fpre[Jb];
}
parameters {
  real<lower=0> kappa;
  real<lower=0> mu;
  real<lower=0> mub[Ja];
  real<lower=0> mua[Jb];
  real<lower=0,upper=1> delta;
  real<lower=0,upper=1> phi;
}
transformed parameters{
  real lambdaa[Ja];
  real lambdab[Jb];
  real kappamu;
  for (i in 1:Jb){
    lambdab[i] <- mub[i]/fpre[i];
  }
  for (i in 1:Ja){
    lambdaa[i] <- delta*mua[i]/fpost[i];
  }
  kappamu <- kappa/mu;
}
model {
  mu ~ gamma(1,0.001);   // prior
  kappa ~ gamma(1,0.7);
  phi ~ beta(1,1);
```

```
  delta ~ beta(1,1);
  mub ~ gamma(kappa,kappamu);   // likelihoods
  mua ~ gamma(kappa,kappamu);
  for (n in 1:Jb) {
    if (ystarbraw[n] == 0)
      increment_log_prob(log_sum_exp(bernoulli_log(1,phi),
      bernoulli_log(0,phi)+poisson_log(ystarbraw[n],lambdab[n])));
    else
      increment_log_prob(bernoulli_log(0,phi) + poisson_log(ystarbraw[n],lambdab[n]));
  }
  for (n in 1:Ja) {
    if (ystararaw[n] == 0)
      increment_log_prob(log_sum_exp(bernoulli_log(1,phi),
      bernoulli_log(0,phi)+poisson_log(ystararaw[n],lambdaa[n])));
    else
      increment_log_prob(bernoulli_log(0,phi) + poisson_log(ystararaw[n],lambdaa[n]));
  }
}
```

Zero-Inflation Bayesian Model for the Paired Design

```
data {
  int J; // number of animals
  int ystararaw[J]; // after treatment McMaster count
  int ystarbraw[J]; // before treatment McMaster count
  int fpre[J];
  int fpost[J];
}
parameters {
  real<lower=0> kappa;
  real<lower=0> mu;
  real<lower=0,upper=1> delta;
  real<lower=0> mub[J];
  real<lower=0,upper=1> phi;
}
transformed parameters{
  real lambdaa[J];
  real lambdab[J];
  real kappamu;
  for (i in 1:J){
    lambdab[i] <- mub[i]/fpre[i];
    lambdaa[i] <- delta*mub[i]/fpost[i];
  }
  kappamu <- kappa/mu;
}
model {
  mu ~ gamma(1,0.001);   // prior
  kappa ~ gamma(1,0.7);
  delta ~ beta(1,1);
  phi ~ beta(1,1);
  mub ~ gamma(kappa,kappamu);   // likelihoods
  for (n in 1:J) {
  if (ystarbraw[n] == 0)
    increment_log_prob(log_sum_exp(bernoulli_log(1,phi), bernoulli_log(0,phi) +
     poisson_log(ystarbraw[n],lambdab[n])));
  else
```

```
    increment_log_prob(bernoulli_log(0,phi) + poisson_log(ystarbraw[n],lambdab[n]));
}
  for (n in 1:J) {
  if (ystararaw[n] == 0)
    increment_log_prob(log_sum_exp(bernoulli_log(1,phi),
     bernoulli_log(0,deltaphi*phi) + poisson_log(ystararaw[n],lambdaa[n])));
  else
    increment_log_prob(bernoulli_log(0,phi) +
     poisson_log(ystararaw[n],lambdaa[n]));

  }
}
```

Appendix B

R Package eggCounts Reference Manual

This appendix contains the reference manual of the R package `eggCounts` version 1.0. The package contains the implementation of the Bayesian hierarchical models that were proposed in this thesis, as well as the improved models from the previous package `eggCounts` version 0.4-1 (Paul, 2015).

Package 'eggCounts'

April 9, 2016

Imports actuar,boot,coda,utils,testthat,numbers

Depends R (>= 3.2.0), Rcpp (>= 0.11.0), rstan, methods

Suggests lattice

Title Hierarchical Modelling of Faecal Egg Counts

Version 1.0

Date 2016-04-09

Author Craig Wang [cre,aut], Michaela Paul [aut], Reinhard Furrer [ctb]

Maintainer Craig Wang <craig.wang@uzh.ch>

Description An implementation of hierarchical models
for faecal egg count data to assess anthelmintic
efficacy. Bayesian inference is done via MCMC sampling in Stan.

License GPL (>= 2)

LinkingTo StanHeaders (>= 2.9.0), rstan (>= 2.9.0-3), BH (>= 1.58.0),
Rcpp (>= 0.11.0), RcppEigen

LazyLoad yes

NeedsCompilation yes

URL http://www.math.uzh.ch/as/index.php?id=eggCounts

RcppModules stan_fit4paired_mod, stan_fit4unpaired_mod,
stan_fit4zipaired_mod, stan_fit4ziunpaired_mod,
stan_fit4nb_mod, stan_fit4zinb_mod

R topics documented:

eggCounts-package *Hierarchical modelling of faecal egg counts*

Description

This package implements Bayesian hierarchical models for the analysis of faecal egg count data. Bayesian inference is done via MCMC sampling in Stan.

Details

Package:	eggCounts
Type:	Package
Version:	1.0
Date:	2016-04-09
License:	GPL (>= 2)
LazyLoad:	yes

Author(s)

Craig Wang <craig.wang@uzh.ch>
Michaela Paul <michaela.paul@uzh.ch>

echinococcus *Faecal egg count sample*

Description

This is an example data set containing 24 eggs per gram (epg) values of Taenia parasites (Echinococcus) in dogs. The correction factor of the diagnostic technique was 50.

Usage

```
data(echinococcus)
```

Examples

```
data(echinococcus)
table(echinococcus)
```

epgs — *Faecal egg count samples (before and after treatment)*

Description

This is an example data set containing 24 eggs per gram (epg) values before and after anthelmintic treatment. The correction factor of the diagnostic technique was 10.

Usage

```
data(epgs)
```

fecrtCI — *Compute standard FECRT according to WAAVP guidelines*

Description

Computes the standard Faecal Egg Count Reduction test together with approximate confidence intervals according to the WAAVP guidelines (Coles et al., 1992, 2006). The function also returns bootstrap percentile confidence intervals.

Usage

```
fecrtCI(epg1, epg2, paired = FALSE, alpha = 0.05, R = 1999)
```

Arguments

epg1	faecal egg counts in untreated animals
epg2	faecal egg counts in treated animals
paired	logical. If true, indicates the samples are paired. Otherwise they are unpaired.
alpha	confidence level of the intervals
R	number of bootstrap replicates

Value

A list with

estimate	the estimated percentage reduction in mean epg rate
bootCI	corresponding percentile bootstrap confidence interval
approxCI	corresponding approximate confidence interval

Author(s)

Michaela Paul <michaela.paul@uzh.ch>

References

Coles GC, Bauer C, Borgsteede FHM, Geerts S, Klei TR, Taylor MA, Waller, PJ (1992). World Association for the Advancement of Veterinary Parasitology (WAAVP) methods for the detection of anthelmintic resistance in nematodes of veterinary importance, Veterinary Parasitology, 44:35-44.

Coles GC, Jackson F, Pomroy WE, Prichard RK, von Samson-Himmelstjerna G, Silvestre A, Taylor MA, Vercruysse J (2006). The detection of anthelmintic resistance in nematodes of veterinary importance, VeterinaryParasitology, 136:167-185.

Examples

```
data(epgs)
fecrtCI(epgs$before, epgs$after, paired=TRUE)
```

`fecr_stan` *Modelling the reduction of faecal egg count data using Stan*

Description

Models the reduction in faecal egg counts data with a (un)paired (zero-inflated) Poisson-gamma model formulation using Stan.

Usage

```
fecr_stan(preFEC, postFEC,rawCounts = FALSE, preCF = 50,
  postCF = preCF, paired = TRUE, zeroInflation=TRUE,
  muPrior, kappaPrior, deltaPrior, phiPrior,
  nsamples = 12000, nburnin = 2000, thinning = 1, nchain = 1,
  ncore = 1, adaptdelta = 0.9, verbose = FALSE)
```

Arguments

`preFEC`	vector of pre-treatment faecal egg counts
`postFEC`	vector of post-treatment faecal egg counts
`rawCounts`	logical. If true, `preFEC` and `postFEC` correspond to raw counts (as counted on the McMaster slide). Otherwise they correspond to calculated epgs (raw counts times correction factor). Defaults to `FALSE`.
`preCF`	correction factor(s) before treatment
`postCF`	correction factor(s) after treatment
`paired`	logical. If true, uses the model for the paired design. Otherwise uses the model for the unpaired design
`zeroInflation`	logical. If true, uses the model with zero-inflation. Otherwise uses the model without zero-inflation
`muPrior`	a list with hyper-prior information for the baseline mean parameter μ. The default prior is `list(priorDist = "gamma",hyperpars=c(1,0.001))`, i.e. a gamma distribution with shape 1 and rate 0.001, its 90% probability mass lies between 51 and 2996

kappaPrior	a list with hyper-prior information for the dispersion parameter κ. The default prior is `list(priorDist = "gamma",hyperpars=c(1,0.7))`, i.e. a gamma distribution with shape 1 and rate 0.7, its 90% probability mass lies between 0.1 and 4.3 with a median of 1
deltaPrior	a list with hyper-prior information for the reduction δ. The default prior is `list(priorDist = "beta",hyperpars=c(1,1))`
phiPrior	a list with hyper-prior information for the zero-inflation parameter ϕ,The default prior is `list(priorDist = "beta",hyperpars=c(1,1))`
nsamples	a positive integer specifying how many iterations for each chain (including burn-in samples)
nburnin	number of burn-in samples
thinning	thinning parameter, a positive integer specifying the period for saving samples
nchain	a positive integer specifying the number of chains
ncore	number of cores to use when executing the chains in parallel
adaptdelta	the target acceptance rate, a value between 0 and 1
verbose	logical. If true, prints progress and debugging information

Details

The first time each model with non-default priors is applied, it can take up to 20 seconds for stan to compile the model. Currently the function only support prior distributions with two parameters. For a complete list of supported priors and their parameterization, please consult the list of distributions in Stan.

Sometimes the function outputs informational message from Stan regarding the Metropolis proposal rejections, this is due to the sampler hit the boundary of the parameter space. For some variables, the boundary point is not supported in the distribution. This is not a concern if there are only a few such warnings.

Value

An object of S4 class `stanfit` representing the fitted results. For more information, please see the `stanfit-class` in `rstan` reference manual.

Prints out the posterior summary of `fecr` as the reduction, `meanEPG.untreated` as the mean faecal egg counts before treatment, and `meanEPG.treated` as the mean faecal egg counts after treatment.

Author(s)

Craig Wang <craig.wang@uzh.ch>

See Also

`simData2s` for simulating faecal egg counts data with two samples

Examples

```
## Not run:
## load the sample data as a vector
data(epgs)

## apply zero-infation model to the data vector
model<-fecr_stan(epgs[,1],epgs[,2],rawCounts=FALSE,preCF=10,paired=TRUE,zeroInflation=TRUE)
```

```
samples<-stan2mcmc(model)

## a demonstration
demo("fecm_stan", package = "eggCounts")

## End(Not run)
```

`fec_stan` *Modelling of faecal egg count data (one-sample case) using Stan*

Description

Models faecal egg counts data in a one-sample case with (zero-inflated) Poisson-gamma model formulation using Stan.

Usage

```
fec_stan(fec, rawCounts = FALSE, CF = 50, zeroInflation = TRUE,
  muPrior, kappaPrior, phiPrior, nsamples = 12000,
  nburnin = 2000, thinning = 1, nchain = 1,
  ncore = 1, adaptdelta = 0.9, verbose = FALSE)
```

Arguments

`fec`	vector of faecal egg counts
`rawCounts`	logical. If true, `preFEC` and `postFEC` correspond to raw counts (as counted on the McMaster slide). Otherwise they correspond to calculated epgs (raw counts times correction factor). Defaults to `FALSE`.
`CF`	correction factor or vector of correction factors
`zeroInflation`	logical. If true, uses the model with zero-inflation. Otherwise uses the model without zero-inflation
`muPrior`	a list with hyper-prior information for the baseline mean parameter μ. The default prior is `list(priorDist = "gamma",hyperpars=c(1,0.001))`, i.e. a gamma distribution with shape 1 and rate 0.001, its 90% probability mass lies between 51 and 2996
`kappaPrior`	a list with hyper-prior information for the dispersion parameter κ. The default prior is `list(priorDist = "gamma",hyperpars=c(1,0.7))`, i.e. a gamma distribution with shape 1 and rate 0.7, its 90% probability mass lies between 0.1 and 4.3 with a median of 1
`phiPrior`	a list with hyper-prior information for zero-inflation parameter. The default prior is `list(priorDist = "beta",hyperpars=c(1,1))`
`nsamples`	a positive integer specifying how many iterations for each chain (including burn-in samples)
`nburnin`	number of burn-in samples
`thinning`	thinning parameter, a positive integer specifying the period for saving samples
`nchain`	a positive integer specifying the number of chains
`ncore`	number of cores to use when executing the chains in parallel
`adaptdelta`	the target acceptance rate, a value between 0 and 1
`verbose`	logical. If true, prints progress and debugging information

Details

The first time each non-default model is applied, it can take up to 20 seconds for stan to compile the model. Currently the function only support prior distributions with two parameters. For a complete list of supported priors and their parameterization, please consult the list of distributions in Stan.

Sometimes the function outputs informational message from Stan regarding the Metropolis proposal rejections, this is due to the sampler hit the boundary of the parameter space. For some variables, the boundary point is not supported in the distribution. This is not a concern if there are only a few such warnings.

Value

An object of S4 class `stanfit` representing the fitted results. For more information, please see the `stanfit-class` in `rstan` reference manual.

Prints out summary of `meanEPG` as the posterior mean egg count.

Author(s)

Craig Wang <craig.wang@uzh.ch>

See Also

`simData1s` for simulating faecal egg count data with one sample

Examples

```
## Not run:
## load the sample data as a vector
data(echinococcus)
fec<-echinococcus[[1]]

## apply zero-infation model to the data vector
model<-fec_stan(fec,rawCounts=FALSE,CF=50,zeroInflation=FALSE)
samples<-stan2mcmc(model)

## a demonstration
demo("fecm_stan", package = "eggCounts")

## End(Not run)
```

`simData1s` *Simulate faecal egg count data (1-sample situation)*

Description

Simulates (zero-inflated) egg count data

Usage

```
simData1s(n = 10, mean = 500, kappa = 0.5, phi = 1, f = 50, rounding = TRUE)
```

Arguments

n	sample size (number of faeces collected)
mean	true number of eggs per gram (epg)
kappa	overdispersion parameter, $\kappa \to \infty$ corresponds to Poisson
phi	prevalence i.e. proportion of infected animals, between 0 and 1
f	correction factor of the egg counting technique, either an integer or a vector of integers with length n
rounding	logical. If true, the Poisson mean for the raw counts is rounded. The rounding applies since the mean epg is frequently reported as an integer value. For more information, please see Details.

Details

The simulation process does not exactly match the proposed models in [ref:paper], however the simulated data resembles the data observed in real world.

In the simulation of raw (master) counts, it follows a Poisson distribution with some mean. The mean is frequently rounded down if it has a very low value and rounding = TRUE, hence there expects to be a negative bias overall when $\mu < 150$. Set rounding = FALSE if one does not wish to have any bias in the simulated counts.

Value

A matrix with three columns, namely the observed epg (obs), number of eggs counted on the McMaster slide (master) and true egg counts (true).

Author(s)

Michaela Paul <michaela.paul@uzh.ch>
Craig Wang <craig.wang@uzh.ch>

See Also

fec_stan for analyzing faecal egg count data with one sample

Examples

```
fec <- simData1s(n=10, mean=500, kappa=0.5, phi=0.7)
```

---

simData2s *Simulate faecal egg count data (2-sample situation)*

Description

Generates two samples of (zero-inflated) egg count data

Usage

```
simData2s(n = 10, preMean = 500, delta = 0.1, kappa = 0.5,
  phiPre = 1, phiPost = phiPre, f = 50, paired = TRUE,
  rounding = TRUE)
```

Arguments

n	sample size (number of faecal samples collected pre- and post-treatment)
preMean	true number of eggs per gram (epg) (i.e. worm burden) before treatment
delta	proportion of epg left after treatment, between 0 and 1. 1 - δ is reduction in mean after treatment, delta = 0.1 indicates a 90% reduction
kappa	overdispersion parameter, $\kappa \to \infty$ corresponds to Poisson
phiPre	pre-treatment prevalence (i.e. proportion of infected animals), between 0 and 1
phiPost	post-treatment prevalence, between 0 and 1
f	correction factor of the egg counting technique, either an integer or a vector of integers with length n
paired	logical. If true, paired samples are simulated. Otherwise unpaired samples are simulated.
rounding	logical. If true, the Poisson mean for the raw counts is rounded. The rounding applies since the mean epg is frequently reported as an integer value. For more information, please see Details.

Details

The simulation process does not exactly match the proposed models in [ref:paper], however the simulated data resembles the data observed in real world.

In the simulation of raw (master) counts, it follows a Poisson distribution with some mean. The mean is frequently rounded down if it has a very low value and rounding = TRUE,there expects to be a up to 3-10% positive bias in the mean reduction when $\mu < 150$ and $\delta < 0.1$. Set rounding = FALSE if one does not wish to have any bias.

Value

A matrix with six columns, namely the observed epg (obs), number of eggs counted on microscope slide (master) and true egg counts (true) for both pre- and post- treatment.

Author(s)

Michaela Paul <michaela.paul@uzh.ch>
Craig Wang <craig.wang@uzh.ch>

See Also

fecr_stan for analyzing faecal egg count data with two samples

Examples

```
fec <- simData2s(n=10, preMean=500, delta=0.1, kappa=0.5)

## show the positive bias when the true reduction should be 95%
set.seed(1)
fec <- simData2s(n=1e5, preMean=150, delta=0.05, kappa=0.5)
1-mean(fec[,5])/mean(fec[,2])
```

stan2mcmc *Convert a Stanfit object to MCMC object*

Description

Converts a large `stanfit` object into a MCMC object for easier analysis, it extracts the relevant MCMC samples of the model from stanFit including the baseline mean epg, dispersion, pre- and post-treatment zero-inflation parameters and the calculated reduction.

Usage

```
stan2mcmc(stanFit)
```

Arguments

`stanFit`	A stanfit object from the output of either `fecr_stan()` or `fec_stan()`

Details

The output can be analyzed as a typical MCMC object with the functions from the `coda` package. NOTE: The resulting MCMC object does not contain warm-up samples and is already thinned.

Value

A MCMC object with a list of relevant parameters depending on the model.

Author(s)

Craig Wang <craig.wang@uzh.ch>

Examples

```
data(epgs)

## apply zero-infation model for the paired design to the data vector
model <- fecr_stan(epgs[,1], epgs[,2], rawCounts=FALSE, preCF=10, paired=TRUE, zeroInflation=TRUE)
samples <- stan2mcmc(model)
summary(samples)
```

tab1morgan *Abundance of trichostrongyloid eggs in sheep faeces*

Description

This data set contains information about the abundance and distribution of trichostrongyloid eggs in the faeces of 14 groups of commercially farmed sheep given in Table 1 in Morgan et al. (2005). The faecal egg counts were assumed to follow a negative binomial distribution with mean m and overdispersion parameter k.

Usage

```
data(tab1morgan)
```

Format

A data frame with 14 rows and 13 variables

Details

The data set has columns:

group	ID number for the groups
ageclass	age class of sheep: "Lambs" or "Ewes"
month	month when samples were taken
n	number of sheep in group
meanFEC	mean number of eggs per gram (epg) of faeces
k	estimated overdispersion parameter k
k.low	lower limit of a 95% confidence interval for k
k.up	lower limit of a 95% confidence interval for k
maxFEC	maximal number of eggs per gram of faeces per group
percentageLarger1000	percentage of samples with more than 1000 epg
Chi2	goodness-of-fit statistic for the negative binomial distribution
df	corresponding degrees of freedom
p	corresponding p-value

Source

Morgan ER, Cavill L, Curry GE, Wood RM, Mitchell ESE (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep, Veterinary Parasitology, 121:79-87.

Examples

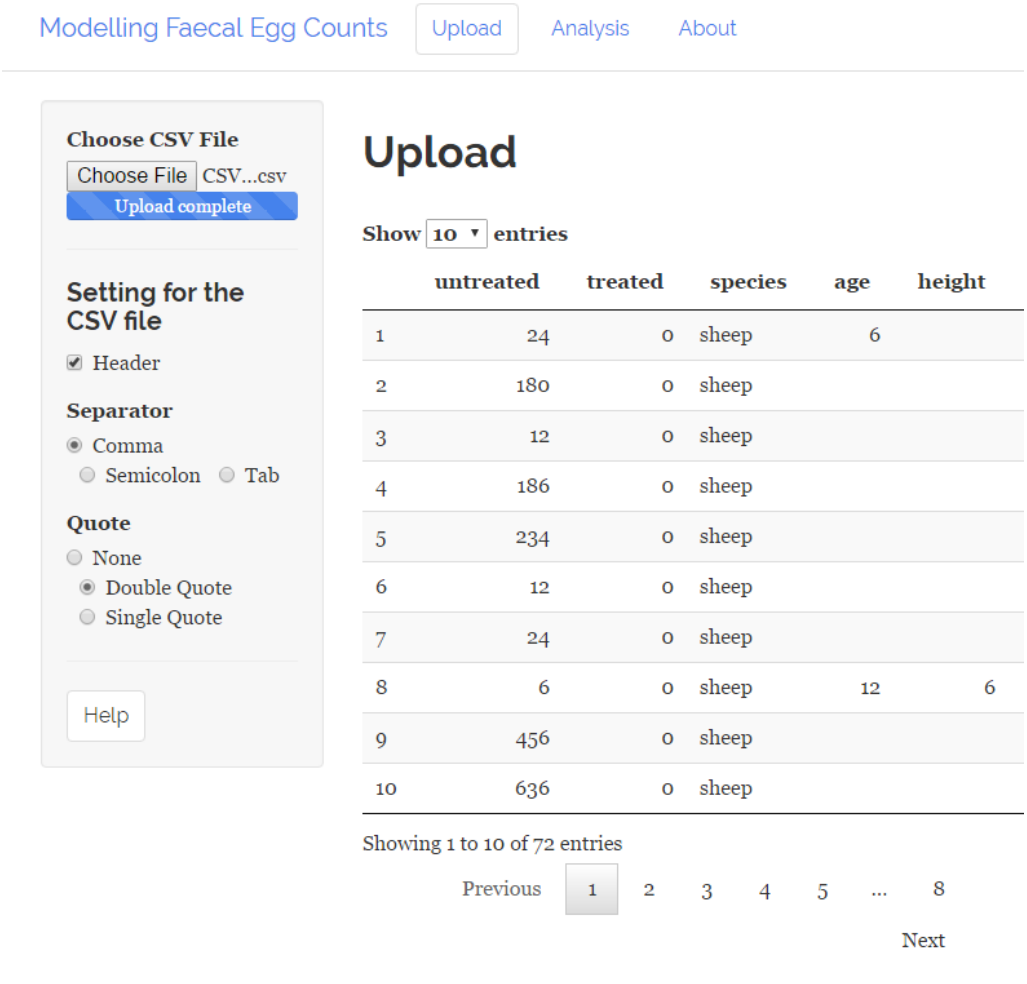
```
data(tab1morgan)
if (require("lattice"))
   xyplot(k.low+k.up+k ~meanFEC, type="p", pch=19, col=c(8,8,1), data=tab1morgan)
```

Index

Appendix C

User-Friendly Web Interface

The user-friendly web interface (Torgerson et al., 2014) written in Shiny (Chang et al., 2016) was based on the previous version of the `eggCounts` package. Extra options and new models have been incorporated into the interface during this thesis.



Modelling Faecal Egg Counts | Upload | Analysis | About

Analysis of

Two Samples Paired

☑ with zero-inflation

FEC (pre treatment)

untreated

FEC (post treatment)

treated

Correction Factor

6

Calculate

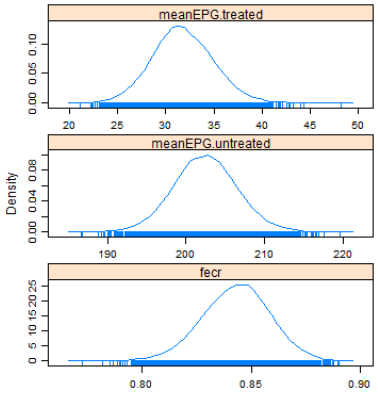
Download FECM Results

Download Descriptive Plot(s)

Download Density Plots

Analysis

	2.5%	50%	97.5%	HPD.Low	mode	HPD.High
fecr	0.811	0.844	0.873	0.812	0.847	0.873
meanEPG.untreated	194.545	202.461	210.550	194.262	202.810	210.169
meanEPG.treated	25.916	31.675	38.078	25.642	31.439	37.708



Modelling Faecal Egg Counts | Upload | Analysis | About

Modelling Faecal Egg Counts with Shiny v1.1

The goal of the shiny-eggCounts project is to provide an intuitive web interface to analyse faecal egg count data. Therefore we developed a Shiny web application whose functionality depends mainly on the R package eggCounts.

For advanced users, we recommend to use the R environment to have access to the full functionality of the package eggCounts.

If you have any suggestions, please send them to Reinhard Furrer .

Authors

Prof. Dr. Reinhard Furrer - Initiator of the shiny-eggCounts project.

Roman Flury - Developed the shiny-eggCounts web application.

Dr. Michaela Paul - Authored the previous version of R package eggCounts

Craig Wang - Improved and re-implemented the models in the R package eggCounts, updated the web application and the current maintainer of the R package eggCounts.

This application was developed at the Institute of Mathematics of the University of Zürich and sponsored by the Swiss National Science Foundation grant CR3313-132482.

Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor .

Title of work (in block letters):

BAYESIAN HIERARCHICAL MODELLING OF ZERO-INFLATED FAECAL EGG COUNTS

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
WANG	CRAIG

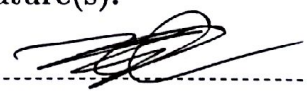
With my signature I confirm that

- I have committed none of the forms of plagiarism described in the Citation etiquette information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work .
- I am aware that the work may be screened electronically for plagiarism.
- I have understood and followed the guidelines in the document *Scientific Works in Mathematics.*

Place, date:

Zürich 10/04/2016

Signature(s):



For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.