Marginal and simultaneous confidence intervals for abundance data with applications to safety assessment of non-target species

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades eines

Doktors der Gartenbauwissenschaften

- Dr. rer. hort. -

genehmigte Dissertation
von

Dipl.-Ing. agr. Frank Schaarschmidt
geboren am 27.02.1979, in Jena.

2009

Referent: Prof. Dr. L. A. Hothorn
Korreferent: Prof. Dr. H.-P. Piepho
Tag der Promotion: 17.12.2008

Abstract

In the approval of novel agricultural practices, inferential statistics can be used to decide between the hazardousness or safety of a novel practice. This might be done based on field trials, which compare a novel treatment to one or several accepted standard treatments and may involve several environments or repeated measurements. A flexible statistical tool for both, summarizing results and allowing decisions, are marginal and simultaneous confidence intervals. Additionally, it might be of interest to include prior knowledge on the parameters of interest into the analysis. While for continous data standard statistical procedures are available, the problem of comparing two or several treatments with respect to the abundance of non-target species is rarely considered.

This work investigates the construction of marginal and simultaneous confidence intervals for ratios and differences of mean abundances in the presence of overdispersion based on the Bayesian framework of Markov Chain Monte Carlo, allowing for the inclusion of prior knowledge. However, main focus is on investigating whether Bayesian intervals constructed can be interpreted as commonly accepted frequentist (simultaneous) confidence intervals when no prior knowledge is available. In simulation studies the coverage probability is assessed. It is found that for such intervals tend to be liberal, but achieve coverage probability close to the nominal level when sample sizes are at least 20 per group or are constructed based on pooled parameters in hierarchical models with larger total number of observations. The nominal coverage is seriously violated, when the samples size is small and the considered species are rare. The application of the methods is shown for two examples from ecological field trials concerning genetically modified crops.

Keywords: inference statistics, negative binomial distribution, multiple comparisons

Zusammenfassung

In der Zulassung von neuer landwirtschaftlicher Verfahren können inferenzstatistische Verfahren genutzt werden über die Bedenklichkeit bzw. Unbedenklichkeit neuer Verfahren zu entscheiden. Grundlage für die Entscheidung können Feldversuche sein, welche eine neue Behandlung mit einer oder mehreren akzeptierten Standardbehandlungen vergleichen und Beobachtungen aus verschiedenen Umwelten oder wiederholte Messungen enthalten. Marginale und simultane Konfidenzintervalle können sowohl zur Zusammenfassung wichtiger statistischer Größen als zur Entscheidung über relevante Hypothesen verwendet werden. Zusätzlich kann es von Interesse sein, Vorwissen bezüglich der betrachteten Parameter in die Analyse einzubeziehen. Während für kontinuierliche Variablen statistische Standardverfahren verfügbar sind, wurde das Vergleiche der mittleren Abundanz von Nichtzielorganismen selten betrachtet.

Die vorliegende Arbeit untersucht die Konstruktion marginaler und simultaner Konfidenzintervalle für Quotienten und Differenzen von mittleren Abundanzen bei vorliegen von Überdispersion auf der Bayesianischen Methode Markov Chain Monte Carlo, die die Einbeziehung von Vorwissen erlaubt. Der Fokus der Arbeit liegt jedoch auf der Frage, ob die resultierenden Intervalle im frequentistischen Sinne interpretiert werden können, wenn kein Vorwissen zugrundeliegt. Zu diesem Zweck wurde die Überdeckungswahrscheinlichkeit in Simulationsstudien untersucht. Darin stellen sich die untersuchten Methoden als liberal dar, erreichen aber ungefähr die geforderte Überdeckungswahrscheinlichkeit, wenn der Stichprobenumfang mindestens 20 beträgt oder die Intervalle für Parameter aus hierarchischen Modellen mit relativ großer Gesamtfallzahl geschätzt werden. Die vorgegebene Überdeckungswahrscheinlichkeit wird grob unterschritten, wenn seltene Spezies auf Basis geringer Stichprobenumfänge untersucht werden. Die Anwendung der diskutierten Methoden wird anhand zweier Beispiele aus ökologischen Feldversuchen mit genetisch veränderten Nutzpflanzen dargestellt.

Schlagworte: Statistik, negative Binomialverteilung, Multiple Vergleiche

Contents

General Notations

The symbols denoted below are used with the same meaning throughout the text. Other symbols may change in their meaning depending on the context and are defined locally.

α	type-I-error probability
N	total number of observations in a statistical model
n	index of the observations $n = 1, ..., N$
I	total number of classes in a classifying variable (of primary interest)
i	index of classes in a classifying variable, $i = 1, ..., I$
M	number of parameters of interest in statistical inference
m	index of the elements of the parameter vector θ, $m = 1, ..., M$
$\boldsymbol{Y}$	$(N \times 1)$ vector of response variable with elements
y_n	elements of $\boldsymbol{Y}$
$\boldsymbol{X}$	$(N \times I)$ design matrix of a linear model
x_{ni}	elements of $\boldsymbol{X}$
$\boldsymbol{C}$	$(M \times I)$ contrast matrix
c_{mi}	elements of $\boldsymbol{C}$
$\boldsymbol{\theta}$	the parameter vector of interest in multiple comparison problems
θ_m	elements of $\boldsymbol{\theta}$, with $m = 1, ..., M$
$\boldsymbol{R}$	$(M \times M)$ correlation matrix in multiple comparison problems
$r_{mm'}$	elements of $\boldsymbol{R}$, with $m = 1, ..., M$, $m' = 1, ..., M$
S	number of simulation runs in a simulation study
K	number of values sampled from the (joint) posterior using MCMC

List of abbreviations in alphabetical order

CI	marginal confidence interval
CPl	coverage probability of a one-sided confidence interval (lower bound calculated)
CPts	coverage probability of a two-sided confidence interval
CPu	coverage probability of a one-sided confidence interval (upper bound calculated)
GM	genetically modified
GMO	genetically modified organism
MCMC	Markov Chain Monte Carlo
pdf	probability density function
SCI	simultaneous confidence intervals
SCS	simultaneous confidence set
SCPl	simultaneous coverage probability of one-sided confidence set (lower bounds)
SCPts	simultaneous coverage probability of a two-sided confidence set
SCPu	simultaneous coverage probability of one-sided confidence set (upper bounds)

Chapter 1

Introduction

1.1 Objectives of safety assessment for novel agricultural practices

Before approving novel agricultural practices, it is of interest to show that they are not harmful for humans, for livestock and, finally, for the various non-target species living in or nearby the crop. The most prominent example for novel agricultural practices is the cultivation of genetically modified (GM) crops. Recently, most public attention has been paid to herbicide (Glyphosate) tolerant crops (Sidhu et al., 2000; Obert et al., 2004; Clark et al., 2006) and to crops which express insecticidal *Cry* proteins (*Bt*-crops) (Berberich et al., 1996; Herman et al., 2004). The following work is motivated by problems which occur when questions on the hazardousness or safety of such new agricultural practices are tackled by inferential statistics following controlled experiments.

The principle of a proof of safety

Assume that inferential statistics shall to be used to come to a decision between the two informal statements: (1) 'The novel practice is harmful' and the (2) 'The novel practice is not harmful'. Then, two statistical hypotheses have to be formulated,

one of which is a falsifiable point-null hypothesis and the other is its complement. Assume further, that the main interest is to directly control the probability of the erroneous decision to conclude for (2) when in fact (1) is true. I.e., main aim is to control the consumers' (or environmental) risk that a harmful agricultural practice is approved. Assume that data y can be obtained which provide information concerning a parameter θ which is a relevant parameter to assess the hazardousness of a novel treatment. Then, statistical hypotheses have to be formulated such that the state of hazardousness of the novel treatment is formulated in the falsifiable null hypothesis H_0, while the state that the novel treatment is not hazardous is formulated in its complement, the alternative hypothesis H_A. For this purpose, one needs to define margins $\underline{\theta}$ and $\overline{\theta}$, marking those values of θ which indicate an unacceptable decrease ($\underline{\theta}$) and an unacceptable increase $\overline{\theta}$.

$$H_0 : \theta \leq \underline{\theta} \cup \theta \geq \overline{\theta} \tag{1.1}$$

$$H_A : \theta > \underline{\theta} \cap \theta < \overline{\theta} \tag{1.2}$$

Based on the intersection-union test principle (Berger, 1982; Casella and Berger, 2002), one can conclude for H_A with error probability α if both elementary null hypotheses $\theta \leq \underline{\theta}$ and $\theta \geq \overline{\theta}$ are rejected by a statistical test with error probability α (e.g. Schuirmann, 1987; Hauschke, 1999; Wellek, 2003). In such a procedure, the consumers' (or environmental) risk has maximally the probability α given that the assumptions underlying the statistical test are correct. Procedures that likewise control the consumers' or environmental risk via the type-I-error probability of a statistical test will be called 'proof of safety' in the following. In the literature, it is also named 'test of equivalence' (Wellek, 2003). The basic problem in practical application of this approach is the definition of $\underline{\theta}$ and $\overline{\theta}$.

In situations, where only an increase or only a decrease leads to concerns about the novel practice, the hypotheses can be defined one-sided (also referred to as 'test on non-inferiority', e.g. Laster and Johnson (2003)). Equations (1.3, 1.4) define a test for non-relevant decrease of θ.

$$H_0 : \theta \leq \underline{\theta} \tag{1.3}$$

$$H_A : \theta > \underline{\theta} \tag{1.4}$$

Practically equivalent, such a test can be performed by using (1.1, 1.2) with $\overline{\theta}$ chosen very large.

The principle of a proof of hazard

However, it is common practice to decide for non-hazardousness of a novel practice if a test of the hypotheses

$$H_0 : \theta = \theta_0 \tag{1.5}$$

$$H_A : \theta < \theta_0 \cup \theta > \theta_0 \tag{1.6}$$

can not be rejected with error probability α, where θ_0 describes a state of non-hazardousness. In such tests, the producers risk is directly controlled via the type-I-error probability α, i.e., the risk of concluding that the novel practice is hazardous if indeed it is not. The consumers' or environmental risk is not directly controlled in such a procedure. However, in advanced applications, statistical methods of power calculation are sometimes used for a post-hoc assessment of the probability to reject (1.5) when the true value of θ is larger or smaller than θ_0 indeed (Andow, 2003; Perry et al., 2003). In the case of a post-hoc power analysis, the practitioner ends up with same the problem as in the proof of safety: to define which changes in θ are considered as relevant changes.

Choosing the parameter of interest

The difficulty of choosing the relevance margins $\underline{\theta}$ and $\overline{\theta}$ changes with the context: When much knowledge is available with respect to the distribution of y in the population, safety margins might be defined directly on the scale of the mean or variance of the observable variable. This might be the case in safety assessment of novel medical treatments or drugs with respect to well known physiological variables y. In other cases, safety margins are defined by law or guidelines of regulatory agencies; an example are the upper limits of GMO impurities in seed lots (Schaarschmidt,

2005). However, the prior knowledge on the distribution of y is usually sparse or it is known that there are major influences on the distribution of y depending on latent covariates. Then, safety of the novel treatment is assessed with respect to a standard treatment, both observed under the same experimental conditions. In this most common case, the parameter of interest θ is a measure of dissimilarity of parameters of the sample exposed to the novel treatment and the sample exposed to one (or several) standard treatment(s). In general settings, both, sample-based comparison between novel and standard(s), and the integration of historical knowledge could be of interest, see, e.g. Berberich et al. (1996); Herman et al. (2004).

When expectations are to be compared between samples, different parameters of dissimilarity are possible: The difference of means is the canonical parameter of dissimilarity when statistical models are assumed to be additive on the scale of y. However, for many biological data this is not a reasonable assumption, since the scale of y is bounded downwardly and/or upwardly. For example, abundance data are defined in $[0; \infty[$. Then, the relevance and even possibility of a difference of $\mu_{Standard} - \mu_{Novum} = 10$ changes dramatically, depending on the particular value of $\mu_{Standard} = 100, 10, 1$. Hence, for strictly non-negative data y, the ratio of means is a more appropriate parameter of dissimilarity in expectations, especially, if the objective is to perform a proof of safety, i.e., showing irrelevant change. Likewise, for data y bounded by a lower and an upper margin, as are binomial data, Wellek (2005) shows that the odds ratio is most appropriate when assessing non-inferiority in comparison to a control, while the difference or ratio of proportions leads to problems in defining proper safety margins independently of the proportion in the control group.

1.2 Relevant distributional assumptions

In the literature, most attention is paid to assessing equivalence for nutritional components, feeding value and other variables concerning humans and livestock, usually being measured on a continuous scale, e.g. vitamin C content or chicken weight, etc.,

(Berberich et al., 1996; Sidhu et al., 2000; Herman et al., 2004; Obert et al., 2004). The usually recommended and applied statistical methods are therefore appropriate for continuous variables (see, e.g. Andow, 2003; Berberich et al., 1996; Sidhu et al., 2000; Herman et al., 2004; Obert et al., 2004). In this work, focus is on safety assessment for non-target species, i.e., on ecological variables measuring the species abundance.

Species abundance is usually measured by counting individuals caught by traps or sweep nets or in visual assessments, in a observational window in time and space. An exception is the collection of plant coverage data (e.g. Warton, 2005) which is not considered here. For counts of particles falling at random into a large number of small areas, which again are distributed at random, or for the number of events occurring at random over a certain period of time can be described by the Poisson distribution (Johnson et al., 1993). A prominent example is the number of particles or rays emitted by a radioactive source and counted by a Geiger counter (Johnson et al., 1993). Obviously, counting insects in certain volumes of air or passing a certain space in a certain period of time can be assumed to result from a similar data generating process. The Poisson distribution assumes that the mean occurrence of the event, μ, is constant over the observations. In this case, the mean and variance of the observed counts have the same value, μ. However, this assumption is frequently violated in ecological data. Then, the variance is higher than the mean and other probability models have to be assumed to account for overdispersion (or 'extra-Poisson variability').

Overdispersion in abundance measurements may occur because of the clustered spatial or temporal distributions of individuals due to behavior or different location of seminal settlement followed by local propagation of individuals. In some cases, overdispersion might be explained by heterogeneity of the observational units with respect to climatic variables or resources. However, in field studies often these variables have not been measured or are not measurable. Even when major predictors for mean abundance are measurable, the population dynamics might lead to relatively large stochastic changes depending on relatively small changes in the variables

(Ives et al., 2008). Even in controlled experiments where major resources and environmental variables are kept constant, overdispersion may be substantial (Beninca et al., 2008). Hence, also models that account for variability among environments by random effects, between repeated measurements by correlation structures or by including covariates to model the effect environmental variables, should still contain a term for overdispersion. For similar reasons, McCullagh and Nelder (1989) recommend the standard use of overdispersion models for count data.

Different probability models have been proposed to describe overdispersed abundance data. First, the quasi-Poisson model assumes a linear dependency of the variance on the mean. The negative binomial distribution, arising from a Gamma-mixture of Poisson variables (Johnson et al., 1993), assumes a quadratic dependency of variance on mean. The applicability of such models to insect abundance data has been controversially discussed: When comparing multivariate species data between different habitats, often zero inflated models are recommended instead of the negative binomial distribution (Welsh et al., 1996; Fletcher et al., 2005; Potts and Elith, 2006). They can account for a higher proportion of zero counts when certain species are completely absent in some habitats and follow a Poisson or negative binomial distribution if present. However, Warton (2005) compares the appropriateness of the negative binomial assumption to other approaches for a number of large abundance data sets, concluding that in most cases the negative binomial assumption fits data best, and is also capable of explaining a high proportion of zero counts. Sileshi (2006) performs a similar comparative study for a number of insect taxa based on the AIC and BIC criterion. He concludes that for the majority of the taxa relatively simple approaches as the negative binomial assumption or the quasi-Poisson method are sufficient.

1.3 Short review of experimental designs in field trials for safety assessment of GMO

In practice, different experimental designs can be found in trials with the aim of safety assessment. Berberich et al. (1996) compare three *Bt* lines with one parental line. Additionally, they compare the observed contents of nutritional and anti-nutritional components to ranges of the components reported from historical observations. Conclusions for equivalence are based on descriptive statistics and one-sided and two-sided paired t-tests.

Sidhu et al. (2000) assess the equivalence of nutritional components in a multi-location trial and an additional feeding study. The multi location trial with 11 sites, treated as random effects, comprised the GM line and a control line as fixed effects. The feeding study was randomized as a randomized complete block design (RCBD) and comprised the GM line and six conventional corn lines. It was analyzed using ANOVA and subsequent t-tests in a mixed model.

Herman et al. (2004) describe a multi location trial to assess equivalence of nutritional components of *Bt* maize. The trial comprised six locations, each being a RCBD with three blocks and three treatments: The GM line, a non-transgenic line and the GM line treated with a herbicide. The conclusion of equivalence is based on a graphical assessment whether observed contents lay within ranges reported from historical trials.

Obert et al. (2004) investigate the equivalence of nutritional components of herbicide tolerant wheat to conventional wheat based on a multi location trial (3 and 5 locations in 1999 and 2000, respectively). Each trial site consisted of a RCBD with four blocks and the GM line and the control line as replicated treatments. Additionally, in each site, four conventional lines were planted without replications at the particular site. However, some lines were also planted in other sites, resulting in a total number of 25 additional conventional lines in the trial. The GM line and the control line were compared in significance tests following a mixed model. The

additional conventional lines were used to calculate tolerance intervals for the nutritional components. These are used subsequently to assess whether the observed values of the GM line fell into these tolerance intervals.

Reynolds et al. (2005) present ranges of nutritional components of maize grain, based on a trial comprising 11 varieties differing in genetic background, usage and climatic adaptation. The trial comprised four geographically distant sites. At each site, seven varieties were planted in an RCBD with three blocks. The data were analysed by a mixed model with variety included as fixed effect and location, block within location and variety-location-interaction treated as random effects. However, purpose of the trial was not comparing GM vs. conventional, but to provide ranges for the contents of various compositional components.

The UK farm scale evaluations compared conventionally grown crops and herbicide tolerant crops. The trial was planned as a multi location trial including replications in five years and considering four different crops. Each location and year comprised a field with two experimental units and to which one replication of each treatment was randomly assigned. This half-field design comprised a total of 272 fields (Perry et al., 2003).

Rauschen et al. (2008) describe a field study on the abundance of planthoppers and leafhoppers exposed to *Bt* maize. The trial was a RCBD with 8 blocks. It comprised three treatments: a maize variety transformed with *Bt* to prevent infestation with the pest *Ostrinia nubilalis*, the variety near-isogenic to the *Bt*-variety without insecticide treatment and the near-isogenic variety with insecticide treatment. The near-isogenic variety and the near-isogenic variety with insecticide treatment clearly serve as two different types of control treatments to assess safety of the *Bt* variety, similar to a negative and a positive control in toxicology.

This limited number of trials concerning safety assessment of GMO may lead to the following considerations: In simple cases, the treatment structure may comprise only two treatments, the GM line and its near isogenic or parental counterpart. However, designs may comprise more than one novel treatment compared to one standard, one

novel treatment compared to several standards, or several standards compared with several novel treatments. If several treatments lead to several elementary hypotheses, the question arises how the elementary hypotheses are connected and whether it is necessary to control the overall (familywise) error rate of the conclusions.

In the field trials, multi-location settings with complex randomization structures dominate, leading to the consideration of mixed models for evaluation of the trials. In more laboratory style experiments, simple randomization structures may also occur.

Often, observed contents of nutritional components are related to the ranges of these components found in previous trials. The information may be found in the literature or databases, e.g., Reynolds et al. (2005). In the considered publications, results are usually presented with focus on the comparison with historical data and the final conclusions are based on the inclusion of the observed values in the historical ranges, rather than significance tests. In some trials (e.g. Obert et al., 2004; Reynolds et al., 2005), considerable effort is spent on assessing the variability of the response variable in the population of standard varieties.

Numbers of replication as small as 3, 4 or 8 on the lowest level of randomization are a common property in the considered experimental designs (Herman et al., 2004; Obert et al., 2004; Rauschen et al., 2008). Occasionally, sample size is not even reported in sufficient detail (Berberich et al., 1996; Sidhu et al., 2000). Only exceptionally, sufficiently large sample sizes are used (Perry et al., 2003).

1.4 Statistical hypotheses for multiple treatment comparisons

A frequently observed setting in safety assessment is the comparison of the novel treatment against several standard treatments. Such designs offer a relaxed definition of safety: Instead of requiring equivalence or non-inferiority to the one and only standard treatment, one may then declare safety if equivalence or non-inferiority can

be shown in comparison to at least one of several standards. Aiming to control the probability of erroneously declaring safety at a low level α, the statistical hypotheses formulating this relaxed safety definition are a union-intersection test (Casella and Berger, 2002). I.e., the global null hypothesis is formed by an intersection of all elementary null hypotheses; the global alternative is a union of the elementary alternatives.

Assume that the dissimilarity between the mean of the novel treatment, μ_1, and the means of the standards, μ_i, $i = 2, ..., I$, is expressed in terms of θ_i, $i = 2, ..., I$, θ_0 indicates the state of equality and decreasing values of $\theta < \theta_0$ indicate hazardousness of the novel treatment, the hypotheses may be:

$$H_0 : \bigcap_{i=2}^{I} \theta_i \leq \underline{\theta}, \tag{1.7}$$

$$H_1 : \bigcup_{i=2}^{I} \theta_i > \underline{\theta}. \tag{1.8}$$

Dunnett and Tamhane (1997) provide solutions for this type of hypotheses by several stepwise test procedures and a single-step procedure assuming a Gaussian error distribution.

Also other combinations of pairwise treatment comparisons may be relevant. As long as the global alternative hypothesis is formulated as the union of all elementary alternatives, simultaneous confidence intervals, as discussed later in this work, are an appropriate statistical tool for deciding on hypothesis controlling the familywise error rate.

A more complicated situation arises, when the aim is to show equivalence of a novel treatment to at least one of several standards, in other words, showing that at least one of several parameters θ_i is between $\underline{\theta}$ and $\overline{\theta}$. Then, the global alternative is a union of intersections of local alternatives. For this situation, alternative methods have been proposed by Bofinger and Bofinger (1993) and Bofinger and Bofinger (1995). The methods considered in this work are not adequate for such situations.

The hypotheses discussed so far are relevant, when the aim is clearly to control the consumers' risk at a low level α. However, when data analysis is still in an ex-

ploratory stage, other hypotheses are adequate: Then, often all pairwise comparisons are performed. When aiming at a proof of hazard involving multiple comparisons, it might be considered fair to protect the producer such that erroneous decisions for hazardousness have a low probability α. In order to extend the scope of this work to such settings, also all pairwise comparisons among multiple treatments, a trend test setting and interaction contrasts are considered.

1.5 Confidence intervals as a concept of statistical inference

Assume that the parameter of interest $\theta \in \Theta$ has been defined. Then, conditional on the sample y and based on a probability model, confidence intervals can be constructed that include the true parameter with pre-specified confidence probability $(1 - \alpha)$. Confidence intervals are not constructed for particular values of θ, as are hypotheses tests. They can be used both as descriptive measures for θ including the uncertainty with respect to θ, and for deciding on hypotheses concerning various values of θ.

As mentioned before, the basic problem of safety assessment is to define margins of relevant change, $\underline{\theta}$ and $\overline{\theta}$, *a priori*. Reasons may be that there is either not enough knowledge for defining such margins, or that there is no consensus about the choice of the margins. In both cases, presenting confidence intervals is the statistical method of choice: In the first case they might merely serve as descriptive measures summarizing effect size, and the associated uncertainty. In the latter case, they summarize all information necessary to perform a proof of safety based on own definitions of relevant change.

Formally, one can reject the null hypothesis of equivalence in Equation (1.1) in favor of the alternative in (1.2) with type-I-error probability α, if and only if a lower $(1 - \alpha)$ confidence bound for θ excludes $\underline{\theta}$ and an upper $(1 - \alpha)$ confidence bound for θ excludes $\overline{\theta}$ (e.g. in Schuirmann, 1987; Hauschke, 1999). As long as

the confidence intervals are constructed as tail intervals, i.e., by inversion of one-sided tests (Agresti, 2003), the above decision rule is technically identical to the inclusion of a two-sided $(1 - 2\alpha)$ confidence interval. The simplification of using $(1-2\alpha)$ confidence intervals for a proof of safety has been criticized (Berger and Hsu, 1996) for particular applications and particular intervals. However, the confidence intervals which are used in the following are either constructed by the analytical inversion of (approximate) tests or are checked by simulations to have the property of (approximate) tail intervals in the sense of Agresti (2003). Hence, their upper and lower bounds should be (approximately) $(1 - \alpha)$ bounds.

In case that one-sided hypotheses as in Equations (1.3) and (1.4) have been defined, one-sided $(1 - \alpha)$ confidence limits can be used for the decision. I.e., if a lower $(1 - \alpha)$ limit for θ excludes $\underline{\theta}$, one can reject the null in favor of the alternative with type-I-error probability α. The validity of this simpler approach has not been subject of discussions comparably to that concerning the two-sided proof of safety.

For deciding on the multiple elementary hypotheses in Equations (1.7) and (1.8) with familywise type-I-error α, simultaneous $(1 - \alpha)$ lower confidence limits for θ_i, $i = 2, ..., I$, can be used. Especially if there is no *a priori* consensus on the choice of $\underline{\theta}$, simultaneous confidence limits summarize all relevant statistical information to enable readers to decide on this type of hypothesis *a posteriori*. When a proof of hazard is the aim, also two-sided simultaneous confidence intervals can be relevant.

1.6 Motivation for Bayesian methods in safety assessment

As outlined above, safety assessment usually takes place in situations ranging between the extremes: (1) complete knowledge concerning the acceptable range of a parameter, obtained from historical experiments or thresholds fixed by conventions; (2) no knowledge concerning the acceptable parameter range of parameters: then, referencing to control treatments is the usual way to define safety margins. As soon

as some prior knowledge is available, it might be included in the statistical analysis via Bayesian methods. In the assessment of equivalence for nutritional components, remarkable effort is spent on obtaining information on the range of variability of the parameters (Obert et al., 2004; Reynolds et al., 2005). Also, much weight is put on comparisons with historical data, when finally drawing the conclusions (Berberich et al., 1996; Herman et al., 2004).

For most biological problems, it is known that the parameter of interest is very likely in a certain subset of the parameter space, but not in the whole real line for continuous, all non-negative integers for counts or all values in the interval $[0, 1]$ for proportions. Hence, the use of weakly informative priors can nearly always be motivated (e.g. Gelman, 2006). For example, when insect abundance is assessed by counting individuals with 10 treatment replications, this implies the experimenters expectation that the species chosen as response variable does not occur with mean abundance greater than say 1000 or less than 0.1. When abundance is greater than 1000 one would probably rather measure biomass instead of individual counts, for abundances less than 0.1 one would either extend the observational window or focus on more abundant species. In this sense, at least 'weakly informative' priors of Bayesian statistics can be motivated. Further, when assessing safety for a particular non-target species, there should be some prior knowledge available, in order to ensure that the species is indeed exposed to the potentially detrimental effects of the novel treatment, as done, e.g. by Rauschen et al. (2008). Hence, some knowledge concerning the ecological properties, behavior and lifestyle of the species must be available for a reasonable choice of the experimental object. Whence such detailed knowledge concerning the species is available, there is usually also some knowledge on the variations of its mean abundance. This prior information might be included in the statistical analysis, at least for the standard treatments.

1.7 Motivation and scope of the following chapters

The above outline of situations and experimental questions in safety assessment for non-target species leads to the focus of statistical investigations in this work:

The distributional assumption needs to account for the special features of overdispersed count data. In the setting of controlled field trials and mesocosm studies, simple assumptions as the negative binomial distribution are appropriate. Zero-inflated models are not considered. In the model that precedes statistical inference, simple one-way layouts as well as hierarchical randomization structures are of interest. Therefore, both, simple fixed effects models as well as hierarchical models are considered. The structure of the treatments which are of primary interest in safety assessment, may comprise only two treatments. Then, marginal confidence limits are sufficient. Frequently, also experiments involving multiple treatment comparisons are designed, making the construction of simultaneous confidence intervals necessary, when the experimental questions are similar to those outlined in Section 1.4. Finally, methods are of interest which, in principle, allow the inclusion of prior knowledge.

The Chapters 3 and 4 briefly introduce some concepts for constructing marginal confidence intervals and simultaneous confidence intervals which are more generally applicable. Chapters 5 and 6 consider the particular problem of marginal and simultaneous confidence intervals of ratios and differences of means of negative binomials in the one-way layout. The focus here is on assessing the frequentist properties of methods derived from MCMC sampling. To expand the scope of multiple comparison procedures, the simulations are not restricted to the most relevant problem of comparisons to control, but also all pairwise comparisons and a trend test scenario is investigated. Chapter 7 provides extensions to simple hierarchical models, again assessing the frequentist simultaneous coverage probability of methods derived from MCMC sampling. The application of the methods to two real data examples is presented in Chapter 8. Finally, Chapters 9 and 10 briefly discuss the main results and outline possible extensions. The main results of the frequentist simulation studies

for particular models are summarized in the Chapters 4, 6, 7, following the description of the models, whereas the `BUGS` code for the models and tables with detailed results can be found in the end these chapters.

Chapter 2

Bayesian inference and Markov Chain Monte Carlo

2.1 Application of the Bayesian Theorem in statistics

Bayes' Theorem considers the probability of an event B given that the event A has been observed, $P(B|A)$. It is generally known in the form

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}, \tag{2.1}$$

where $P(A)$ denotes the probability of observing event A. For simplicity, consider a setting where an observable random variable Y is thought to emerge according to a certain probability model in dependence of a parameter θ. Applying Bayes' Theorem with the aim of statistical inference, allows a statement concerning the probability of a parameter θ, given the observed data y (Ellison, 2004):

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)}. \tag{2.2}$$

Note, that θ is understood as (unobservable) random quantity. $P(y) = \int_\theta (P(y|\theta)\,P(\theta))\,d\theta$ is the marginal density of Y resulting from integrating over all possible values of θ. In practice, the probability to observe data y given a parameter θ, $P(y|\theta)$, follows from the distributional assumption and the prior knowledge

concerning θ is summarized in a density $P(\theta)$. Then, Equation (2.3) is the complete probability model connecting the observable y, the unobservable parameter θ, and the prior knowledge (Gilks et al., 1996).

$$P(y, \theta) = P(y|\theta)\, P(\theta) \tag{2.3}$$

However, main interest is in a generalizable statement concerning θ, given the prior information and the observed data. This is summarized by the posterior distribution, as defined in Equation (2.4) for general cases (Gilks et al., 1996):

$$P(\theta|y) = \frac{P(y|\theta)\, P(\theta)}{\int P(y|\theta)\, P(\theta)\, d\theta} \tag{2.4}$$

In simple cases, e.g. the one parameter binomial probability model, the distribution $P(\theta|y)$ can be described analytically (Gelman et al., 2004). In most cases of practical interest, Markov Chain Monte Carlo (MCMC) methods have to be used to draw sufficiently large samples from $P(\theta|y)$ to approximate the distribution. Technical background of MCMC in general and the Gibbs sampler in particular are outlined in Section 2.6.

2.2 Bayesian vs. frequentist inference

Bayesian and frequentist methods differ in the assumption concerning the parameter: Frequentist approaches assume parameters to be unknown but fixed quantities, while in Bayesian analysis, parameters are considered as random quantities which are unobservable and are contrasted to the observable random quantities which are commonly called variables. In that sense, the differentiation between random and fixed parameters in frequentist mixed models is less clear in the Bayesian context (Gelman et al., 2004).

Confidence intervals provide a range of parameter values, in which the parameter of interest lies with a pre-specified confidence probability $(1 - \alpha)$. Bayesian and frequentist approaches differ in the interpretation of this probability: Frequentist confidence intervals are designed such that the true, fixed parameter lies within the

confidence bounds in $(1-\alpha)100\%$ of the cases when a given random experiment is repeated $n \to \infty$ times (e.g. Ellison, 2004). It is not correct to state, for a given realization of the random experiment, that the true parameter lies within the confidence interval with probability $(1-\alpha)$. Rather, it is still unknown whether the true value lies within the confidence interval; it is only known that it would lie within $(1-\alpha)100\%$ of the intervals computed by the same method, when the experiment would be repeated many times. In Bayesian analysis, the somewhat unpleasant frequentists interpretation of probability is replaced by a subjective interpretation of probability. I.e., probability is interpreted as the 'subjective belief in the probability of an event' (Ellison, 2004): The beliefs concerning the distribution of the parameter which are known prior to the experiment are summarized in the prior distribution. The posterior distribution, i.e., the prior belief, updated by the information concerning the parameter of interest obtained from the sample, is still subjective. The $(1-\alpha)$ confidence bounds obtained from this distribution can properly be interpreted as a range of parameter values which are believed to occur with probability $(1-\alpha)$ (Ellison, 2004). For that reason, 'Bayesian confidence intervals' (Carlin et al., 2006) are commonly referred to as 'credibility intervals' or 'credible intervals'.

2.3 Choice of prior distributions

A prior density is called proper, if it is independent of the data and integrates to 1 (Gelman et al., 2004). To ensure that prior distributions are proper is important for the following inference based on the posterior distribution. However, even an improper prior might lead to a proper prior distribution and therefore valid inference (Carlin et al., 2006). Nevertheless, this has to be checked in each particular case (Gelman et al., 2004) and therefore improper priors will not be used here.

A prior is called conjugate, if the posterior density $p(\theta|y)$ following from the combination of the sampling distribution $p(y|\theta)$ and the prior $p(\theta)$ has the same parametric form as the prior. This can be shown analytically for a number of simple models, e.g. for assuming a one parameter binomial sampling distribution and a

Beta prior distribution (Gelman et al., 2004). Choosing the prior to be conjugate allows analytical solutions for the posterior in simple cases, and to interpret the prior information as additional data (Gelman et al., 2004). However, if a chosen prior is not conjugate, this does not lead to problems in the inference based on the posterior distribution (Gelman et al., 2004). Also, for complex models, conjugate priors may not even exist (Gelman et al., 2004). Hence, whether a prior is chosen conjugate or not, is not discussed here in detail for the particular models, but priors are chosen similar to those found in related models in the literature and worked examples.

Priors for mean and variance parameters

Gelman et al. (2004) recommend not to use non-informative priors when the number of parameters is large, but to use hierarchical models with hyperpriors instead. In the generalized linear model, imposing a normal prior on the mean parameter β on the scale of the link is usually a non-conjugate prior (Gelman et al., 2004).

When choosing a non-informative prior for variance parameters (or inverse variance parameters), distributions are needed the cover the range $[0, \infty]$. A natural candidate is the gamma distribution. However, the adequate choice of priors for variance parameters is under discussion, especially for complex models (Zhao et al., 2006; Browne and Draper, 2005; Gelman, 2006; Kass and Natarajan, 2006). While uniform, gamma and several folded distributions are discussed, the uniform and gamma distribution are commonly found in worked examples (Zhao et al., 2006; Spiegelhalter et al., 2007). Gelman et al. (2004) notes that it is hard to define principles for the choice of prior distributions that are appropriate in all cases. For example, although intuitively reasonable, uniform priors should be used with care for precision parameters on the log scale or in mixture models (Gelman et al., 2004). In this work, only gamma and uniform priors are used for variance parameters.

Non-informative priors

Generally, if there is no information concerning θ, the prior distribution $P(\theta)$ has a flat (vague, diffuse, about uniform) distribution in the parameter space Θ. Then, it contributes an about equal, very small probability to Equation (2.4), independent of the particular choice of θ. In this case, the posterior depends only on the sample y and the imposed distributional assumption. However, if Θ is unbounded, i.e., $\Theta = [0, \infty]$ or $\Theta = [-\infty, \infty]$, a truly uniform prior that is also a proper prior is not possible. Occasionally, further principles are applied when choosing a non-informative prior density: Jeffreys priors are chosen such that the result of the analysis is invariant with regard to a transformation of the parameter of interest; this is most important in the case of the binomial parameter, where invariance with regard to the exchange of success and failure is often required (Gelman et al., 2004; Agresti and Min, 2005). When non-informative priors are assumed for the parameters, Bayesian analysis approaches the solutions found by frequentist analysis (Gelman et al., 2004).

The prior $\theta \sim N(0, 1000)$ will in the following be used for mean parameters with $\Theta =]-\infty, \infty[$. The priors $\theta \sim gamma(0.001, 1000)$ or $\theta \sim unif(0, 1000)$ will be used for variance parameters with $\Theta = [0, \infty[$ in the following considerations. Although both choices are frequently found in worked examples (Zhao et al., 2006; Spiegelhalter et al., 2007), Zhao et al. (2006) recommends the use of gamma priors only with smaller variance, e.g. $\theta \sim gamma(0.01, 100)$. Gelman (2006) discourages the use of the family of gamma priors because this distribution is not truly non-informative and the resulting inference depends on the particular choice of the parameters a and b of the gamma distribution (See Appendix A for the parametrization used here). The discussion among Browne and Draper (2005); Gelman (2006); Zhao et al. (2006) shows that there is no consensus on this issue.

2.4 Safety assessment in the Bayesian context

Purely Bayesian approaches to assess equivalence of novel treatments to a standard treatment often use different approaches to decide between the two informal statements (1) 'The novel practice is harmful' and the (2) 'The novel practice is not harmful'. Most interesting is the concept of a loss function as discussed by Lindley (1998) as an alternative to the definition of fixed margins $\underline{\theta}, \overline{\theta}$. Other authors explicitly test hypotheses of equivalence as defined in Equations (1.1) and (1.2), using decision rules based on Bayes factors (Williamson, 2006) and posterior probabilities of hypotheses (Selwyn et al., 1981; Ghosh and Rosner, 2007). None of these approaches aims at controlling the type-I error at a low pre-specified level. In this work, such alternative approaches to decide for equivalence of two treatments are not considered.

2.5 Multiple comparisons in the Bayesian context

Various authors discuss the problem of multiple comparisons in the Bayesian context, with focus on problems of multiple hypotheses testing, with elementary null hypotheses describing the state of no effect (Westfall et al., 1997; Berry and Hochberg, 1999; Efron et al., 2001; Chen and Sarkar, 2004; Scott and Berger, 2006; Ji et al., 2008; Pennello, 2007). This is motivated by the important field of application in microarray experiments (e.g. Efron et al., 2001; Efron, 2004; Scott and Berger, 2006; Efron, 2007; Ji et al., 2008) with focus on criteria similar to the false discovery rate rather than familywise error rate, and no focus on interval estimation.

Procedures for multiple hypotheses testing in the Bayesian context usually differ from frequentist procedures in both, the aims and approaches. Starting with Duncan (1965) and Waller and Duncan (1969) (reviewed in Berry and Hochberg, 1999), procedures are proposed which decide between null and alternative hypothesis by loss functions that control the ratio of type-I to type-II errors (e.g. Scott and Berger, 2006; Pennello, 2007). Hence, such procedures usually do not control the family-

wise error rate and do not aim to do so. Other approaches try to avoid finding too many effects in multiple hypothesis testing by modelling the parameters of interest as members of a population, which is modeled by common hyper parameters. I.e., the parameter of interest which are tested in the elementary hypotheses are jointly modeled as a random effect. In this way, estimates are shrunk towards each other, making the analysis more conservative (Berry and Hochberg, 1999). Using Dirichlet priors for jointly modelling the parameters of interest allows to calculate posterior probabilities of specific hypotheses, e.g. for the hypotheses of all pair wise comparisons (Berry and Hochberg, 1999). However, such models usually rely on the assumption that the single parameters are exchangeable. This assumption appears acceptable in the context of microarrays but not in the context discussed in this work, where the parameters of interest are known to be correlated in non-trivial structures (see Chapter 4). Only few papers explicitly consider multiple treatment comparisons, which are of main interest in this work. Examples are Westfall et al. (1997) and Berry and Hochberg (1999) for all pairwise comparisons, Chen and Sarkar (2004) for comparisons to control and Nashimoto and Wright (2008); Shang et al. (2008) for appraoches to include order restriction. Simultaneous intervals are not considered by these authors.

A common Bayesian argument against multiplicity adjustments is that controlling the familywise error rate is only important when many or all null hypotheses could be plausibly true. Conversely, when there is relatively firm prior knowledge indicating that some of the null-hypotheses are true, some are false, the necessity to adjust for multiple comparisons in order to control the familywise error rate decreases (Westfall and Young, 1993). Westfall et al. (1997) investigate this point more closely. They find that, depending on the prior probability that is assigned to the global null hypothesis in multiple testing, Bayesian reasoning leads to similar adjustments as that of Bonferroni for independent hypotheses or a frequentist adjustment proposed for all pairwise comparisons. However, in safety assessment for non-target species, firm prior knowledge concerning the safety hypotheses is rarely available. The reasons are that sampling is expensive and there is relatively low public interest in non-target

species compared to safety or efficacy analyses in clinical trials; hence, prior knowledge is usually sparse. Prior knowledge may be reasonably available for the range of mean abundances in conventional treatments, but will be rarely available for the novel treatments. Hence, when primary interest is in parameters of dissimilarity between standard and novel treatments, there is usually no firm prior knowledge available, and the above arguments against multiplicity corrections do not apply.

2.6 Short introduction to Markov Chain Monte Carlo (MCMC)

MCMC aims to draw samples from the posterior density of the parameter of interest θ conditional on the sample y: $p(\theta|y)$.

1. T updates $\theta^{(t)}$, $t = 1, ..., T$, are drawn, where $\theta^{(t)}$ depends only on $\theta^{(t-1)}$,
2. at each step t, a value θ' is drawn based on the density centered at $\theta^{(t)}$.
3. The parameter value is updated, $\theta^{(t+1)} = \theta'$, or the last value is kept $\theta^{(t+1)} = \theta^{(t)}$, depending on ratios $p\left(\theta'|y\right) / p\left(\theta^{(t)}|y\right)$ and further transition probabilities depending on $\theta^{(t)}$ and θ',
4. by frequent updates, the distribution of sampled values $\theta^{(t)}$ converges to $p(\theta|y)$.

Different available algorithms differ in the way by which transition probabilities are calculated for the decision whether the parameter is updated $\theta^{(t+1)} = \theta'$ or not $\theta^{(t+1)} = \theta^{(t)}$ (Gelman et al., 2004; Congdon, 2006).

The Gibbs sampler is especially designed for multi-parameter problems. It might be used with different algorithms (but all using the rejection sampling strategy above) to perform the single steps in the following structure:

- The vector θ is split into D sub vectors $(\theta_1, ..., \theta_D)$, $d = 1, ..., D$.
- In the tth update of a particular subvector $\theta_{d'}$, $\theta_{d'}^{(t)}$ is updated depending on $\theta_{d'}^{(t-1)}$ for $d' = d$ but depending on $\theta_d^{(t)}$ for $d \neq d'$

(Gelman et al., 2004). The chain $\left(\theta^{(1)}, ..., \theta^{(T)}\right)$, $t = 1, ..., T$ sampled in such a way, starts at initial values and converges to the target distribution $P(\theta|y)$ with increasing t. Hence, the part of $\theta^{(1)}, ..., \theta^{(T)}$ with sufficiently high t can be taken as a sample from $P(\theta|y)$. In the following, such a sample will be denoted $\boldsymbol{\theta}^*$

2.7 Assessing the frequentist performance of MCMC based confidence intervals

As outlined in the introduction, safety assessment might need both, frequentist or Bayesian approaches, depending on the availability of prior information and the acceptance of Bayesian methods. In the Bayesian context, analysis with non-informative priors might be presented to 'let the data speak for themselves' (Gelman et al., 2004). In this case, the posterior depends only on the observed sample, and therefore is not influenced by subjective information except the choice of the distributional assumption. That is the reason why occasionally non-informative priors are called objective. A statistical procedure might then be validated by its frequentist performance. Additionally, for situations with more complex designs and unusual distributional assumptions, simple frequentist methods may provide unacceptable solutions (an example is given in Chapter 5) and more appropriate frequentist approaches have so far not been generalized to cover particular problems. In these situations, the Gibbs sampler is a flexible tool to sample from 'objective' posterior distributions for the parameters of interest. Whether confidence intervals obtained in this way may be interpreted in a frequentist manner can then be assessed by simulation studies.

Numerous authors assess the matching of non-informative Bayesian and frequentist solutions for special problems, either by analytical approaches (Welch and Peers, 1963; Peers, 1965; Nicolaou, 1993; Datta and Ghosh, 1995; Datta, 1996; Sweeting, 2001; Ghosh et al., 2003; Dilba, 2006; Berger and Sun, 2008, to name some) or by simulation studies or other approaches to describe the empirical properties (e.g.

Nicolaou, 1993; Ghosh and Kim, 2001; Agresti and Min, 2005; Shi and Bai, 2008). Some analytic approaches which provide solutions for particular problems focus on relatively simple problems as estimating various parameters in the bivariate normal model (Berger and Sun, 2008), estimation for the two-parameter gamma distribution (Nicolaou, 1993), parameters following a general linear model (i.e. Gaussian error distribution, Ghosh et al. (2003)), models for 2×2 contingency tables (Agresti and Min, 2005; Shi and Bai, 2008) or the comparisons of two means in the Behrens-Fisher problem (Ghosh and Kim, 2001). More complex problems with computations based on the Gibbs sampler have not been addressed so far. Also the specific problem of frequentist performance of simultaneous credible sets for contrasts of mean parameters in generalized linear models has not been considered. In the remaining part of this work, the main focus is on the question, whether marginal and simultaneous credible intervals based on samples of the joint distribution derived by the Gibbs sampler lead to valid frequentist marginal and simultaneous confidence intervals, when non-informative priors are used and interest is in various models for count data. Since for a part of the considered models and inferential problems, frequentist methods have not been proposed so far, the frequentist coverage probabilities are assessed by simulation studies.

Definition of coverage probabilities

Assume a confidence interval $\left[\hat{\theta}_L, \hat{\theta}_U\right]$ for the parameter θ. The coverage probability of a two-sided confidence interval (CPts) is then defined $P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U\right)$ when the random experiment of calculating the confidence interval based on random data is performed many times. Accordingly, the coverage probability of the one-sided lower confidence interval (CPl) is then defined as $P\left(\hat{\theta}_L \leq \theta\right)$, and the coverage probability of the one-sided upper confidence interval (CPu) is defined as $P\left(\theta \leq \hat{\theta}_U\right)$.

Assume that $\left[\hat{\theta}_{mL}, \hat{\theta}_{mU}\right]$, $m = 1, ..., M$ define the bounds of the confidence set for an M-dimensional parameter $\boldsymbol{\theta}$. The simultaneous coverage probability of a two-sided confidence interval (SCPts) is defined as $P(\hat{\theta}_{mL} \leq \theta_m \leq \hat{\theta}_{mU}, \forall m = 1, ..., M)$. Accordingly, SCPl is defined as $P(\hat{\theta}_{mL} \leq \theta_m, \forall m = 1, ..., M)$ and SCPu is defined

as $P(\theta_m \leq \hat{\theta}_{mU}, \forall m = 1, ..., M)$.

Technical details

For the simulation study, `R-2.6.0` (R Development Core Team, 2007) has been used. The Gibbs sampler implementation `OpenBUGS 3.0.2` (Spiegelhalter et al., 2007) has been used to draw samples from joint posterior distributions. For calling `OpenBUGS` from within `R`, the package `R2WinBUGS 2.1-6` (Sturtz et al., 2005) has been used.

The `BUGS` language allows flexible definition of various models and also mis-specified models can be run as long as they are syntactically correct and fitting data and initial values are provided. Hence, for completeness and reproducibility, the `BUGS` code used in the simulations is always shown next to the simulation results. Attached to the `BUGS` code are technical details for the MCMC updates, since the convergence and hence the quality of representation of the target distribution by the MCMC sample may greatly depend on these details and the particular `BUGS` code.

Quality of approximation of the (joint) posterior distribution

The quality of MCMC derived intervals can only be sufficient, when they are constructed based on a sufficient number of values K, truly sampled using MCMC from the posterior distribution of interest. Both, a small sample K used for interval construction and sampling from a distribution that has not readily converged to the posterior of interest will lead to credible intervals with insufficient properties. The convergence of the sampling distribution to the distribution of interest crucially depends on the combination of the initial values, the number of updates, the autocorrelation between consecutive updates, the number of observations discarded at the beginning of the chain of updates (burn-in) and finally, the frequency by which sampled values are discarded within the sampled chain (thinning):

1. The initial values are 0 for parameters with $\theta \in [-\infty; \infty]$ and for 1 for parameters with $\theta \in [0; \infty]$. This is clearly a suboptimal choice: In real world

applications, convergence might be sped up by using sample estimates as initial values. For all simulations, only one MCMC chain has been run.

2. The number of updates finally used for interval construction, is usually K=1000.

3. For the considered models and their `BUGS` implementations, the autocorrelation increases with complexity of the models.

4. The number of updates discarded at the beginning of the chain (burn-in) needs to be increased when autocorrelation slows down the convergence. The burn-in has been chosen specifically for each model in particular based on simulated example data, such that Gewekes test on convergence (Gelman et al., 2004; Geweke, 1992) did not show deviations.

5. The number of values discarded within the chains was chosen for each model specifically based on simulated example data, such that autocorrelation plots in `OpenBUGS` or the function `acf` in the R-package `coda` did not show autocorrelations for lags greater or equal 2.

Due to the high computational intensity of MCMC sampling for complex models, the number of random samples $\boldsymbol{y}$ drawn from the assumed distribution in order to assess the coverage probability of the intervals is usually S=1000. Repeating a binomial experiment 1000 times, the probability of success π (here defined as the probability to cover the true parameter) such that the hypothesis $H_0 : \pi \geq 0.95$ can be rejected at the 0.05 level for less than 936 out of 1000 events.

Chapter 3

Concepts for constructing confidence intervals

In this Chapter, general concepts to construct confidence intervals for a parameter θ are shortly reviewed.

3.1 Desirable properties of confidence intervals

Let θ be the parameter of interest in parameter space Θ, and denote $\left[\hat{\theta}_L; \hat{\theta}_U\right]$ the confidence interval for θ derived from sample y.

1. In the frequentist setting, the interval should cover θ with probability $(1-\alpha)$.
2. Two-sided confidence interval should be symmetric in the probability to exclude the true parameter θ, i.e. equi-tailed: When a two-sided $(1-2\alpha)$ confidence interval is constructed, the probability to exclude θ should be α for both the lower and the upper bound.
3. The confidence intervals should be informative when the sample y is informative with respect to θ, i.e., $\left[\hat{\theta}_L; \hat{\theta}_U\right] \subset \Theta$, and not $\left[\hat{\theta}_L; \hat{\theta}_U\right] \subseteq \Theta$!
4. The confidence interval should be non-degenerate for incomplete knowledge with respect to θ and reasonable outcomes Y, i.e., $\hat{\theta}_L < \hat{\theta}_U$.

5. If the evidence for $\theta > \theta_0$ is higher in sample y_i than in sample y_j, it is required that $\hat{\theta}_{Li} > \hat{\theta}_{Lj}$; correspondingly, if the evidence for $\theta < \theta_0$ is higher based on sample y_i than sample y_j, it is required $\hat{\theta}_{Ui} < \hat{\theta}_{Uj}$.

3.2 Wald-type confidence intervals based on normal approximation

When it can be shown that for large N the distribution of $T = \frac{\hat{\theta}-E(\theta)}{\sqrt{\hat{V}(\hat{\theta})}} \sim N(0,1)$, a Wald-type confidence interval can be constructed by analytically inverting the (Wald-type) test by solving for θ_0 in Equation (3.1)

$$\frac{\hat{\theta} - \theta_0}{\sqrt{\hat{V}\left(\hat{\theta}\right)}} = z. \tag{3.1}$$

These confidence limits then are:

$$\left[\hat{\theta}_L; \hat{\theta}_U\right] = \left[\hat{\theta} \pm z\sqrt{\hat{V}\left(\hat{\theta}\right)}\right]. \tag{3.2}$$

Here, $z = z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. When the normal approximation is sufficiently precise, such intervals achieve the objectives stated in Section 4.1. When the normal approximation is inappropriate due to small sample sizes, discrete, skewed, or kurtotic distributions of $\hat{\theta}$ at least some objectives in Section 4.1 are violated.

The term Wald-type, e.g. in Brown et al. (2001); Brown and Li (2005), refers to the fact that the variance of $\hat{\theta}$, $\hat{V}\left(\hat{\theta}_m\right)$ contributes as a constant to the Equations (3.1) and (3.2). The uncertainty that is introduced in the dependency of $V\left(\hat{\theta}\right)$ on θ by the uncertainty on θ itself is not taken into account. However, this procedure is reasonable for large sample sizes, most simple, applicable for many scenarios and can be easily extended to the problem of multiple comparisons (Chapter 4). However, for small sample sizes and discrete distributions with parameters close to the boundary of the parameter space, obtained variance estimates can be unreasonable (e.g. Brown et al., 2001). Then, single events may exist that result in confidence intervals which

are degenerated to a point or comprise the whole parameter space. Hence, they may violate the desirable properties outlined in 3., 4., and 5. in Section 3.1 above.

3.3 Profile likelihood methods to construct confidence intervals

The latter problem of Wald-type intervals for parameters of discrete distributions may be overcome by intervals that are constructed by point-wise inversion of a test for a sufficient number of points $\theta_0 \in \Theta$, and which include the variance of $\hat{\theta}$ based on θ_0 rather than $\hat{\theta}$. For simple problems, such tests can even be inverted analytically (Wilson, 1927). More generally applicable concepts are the point-wise inversion of likelihood ratio tests, e.g. Venzon and Moolgavkar (1988) or the construction of intervals based on deviance profiles (Chen and Jennrich, 1996). For example, confidence intervals based on the inversion of a likelihood ratio test (Chen and Jennrich, 1996) can be constructed by evaluating Equation (3.3) for a sufficient number of values for θ_c. Let $L(\theta)$ denote the likelihood function derived from the probability model assumed for the creation of Y given θ. Then a confidence interval based on the profile likelihood method is given by

$$I = \left\{ \theta_c : -2\log\frac{L(\theta_c)}{L\left(\hat{\theta}\right)} \leq \chi^2_{1,1-\alpha} \right\}, \tag{3.3}$$

where $\chi^2_{1,1-\alpha}$ is the $1-\alpha$ quantile of the χ^2 distribution with 1 degree of freedom and I the set of values which forms the confidence set, with its smallest and largest values $[\hat{\theta}_L; \hat{\theta}_U]$. Instead of comparing the likelihood ratio vs. a χ^2 quantile, other functions $f(\theta|y)$, e.g. the deviance, might be compared vs. quantiles of other adequate distributions.

Constructing confidence intervals based on this approach is computationally more intensive and expanding its idea to construct simultaneous confidence intervals including the correlation among parameters (see Chapter 4 for a brief motivation) is not straightforward. Of course, simultaneous confidence intervals based on profile

methods can be constructed using the Bonferroni correction.

3.4 Confidence intervals based on an empirical (posterior) distribution

Assume a sample $\boldsymbol{\theta}^* = (\theta_1^*, ..., \theta_K^*)$ of K values from the posterior distribution $P(\theta|y)$ has been drawn using an MCMC implementation, where the values are indexed by $k = 1, ..., K$. Then, a $(1-\alpha)$ credible interval for θ can be derived by calculating the sample quantiles $\hat{Q}_{\alpha/2}$ and $\hat{Q}_{1-\alpha/2}$ from $\boldsymbol{\theta}^*$. Hyndman and Fan (1996) summarize the properties of several definitions to calculate $\hat{Q}_\alpha$. Here, definition 7 of Hyndman and Fan (1996) is used. The obtained values divide the range of $\boldsymbol{\theta}^*$ into $k-1$ intervals, of which 100α % lie to the left of $\hat{Q}_\alpha$ and $100(1-\alpha)$ % lie to the right of $\hat{Q}_\alpha$.

Intervals constructed in such a way should have the properties of tail intervals (property 2 in Section 3.1). Such intervals are favored since interest is in both, estimation and hypothesis testing with subsequent one-sided interpretation. However, in case that interest would be purely in estimation, also highest posterior density (HPD) intervals (Gelman et al., 2004; Geweke, 2005) would be an appropriate method. These do not have the property of a tail interval for skewed posteriors (Agresti and Min, 2005), and hence are not considered in this work.

Chapter 4

Concepts for constructing simultaneous confidence intervals

In this Chapter, general concepts to construct simultaneous confidence intervals (SCI) for an M-dimensional parameter vector $\boldsymbol{\theta}$, with elements θ_m, $m = 1, ..., M$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, are reviewed. As a first step one might consider the problem of constructing a simultaneous confidence set for $\boldsymbol{\theta}$, i.e., a subspace of $\boldsymbol{\Theta}$ in which the true parameter vector $\boldsymbol{\theta}$ is included with probability $(1 - \alpha)$. Such sets might be constructed by pointwise inversion of a simultaneous test and could have shapes very different from a hyper rectangle (Dilba, 2005). However, for interpretations in problems with $M > 2$, usually a projection to the axes is needed, resulting in simultaneous confidence limits that form a hyper rectangle in $\boldsymbol{\Theta}$. For this reason, in this work only such methods are considered which lead to rectangular confidence sets; these are referred to as simultaneous confidence intervals (SCI) for simplicity.

4.1 Desirable properties of simultaneous confidence intervals

Additional to the objectives defined in Section 3.1, one may define the following properties of simultaneous confidence intervals as desirable:

1. Simultaneous coverage probability: Cover $\boldsymbol{\theta}$ with probability $(1-\alpha)$. For am M-dimensional confidence set with bounds $\left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right]$ for its mth element, the simultaneous coverage probability is formally defined as $P\left(\theta_m \in \left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right];\forall m=1,...,M\right)$.
2. When used to obtain decisions on hypotheses, confidence sets should be equitailed, i.e., when a two-sided $(1-\alpha)$ confidence set is constructed by $2M$ limits, the probability to exclude a part of $\boldsymbol{\theta}$ by each bound should be $\alpha/(2M)$.

4.2 Simple solutions: Bonferroni and Sidak

Objective 1 in Section 4.1 can be best achieved when the correlation among the M parameters of interest is taken into account. However, simple solutions exist that perform conservative in common situations; the term conservative is used for the situation $P\left(\theta_m \in \left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right];\forall m=1,...,M\right) > 1-\alpha$.

Bonferroni adjustment

Motivated by Equation (4.1) below, the Bonferroni adjustment preserves the simultaneous coverage probability by constructing marginal confidence intervals $\left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right]$ with level $(1-\alpha/M)$ for each θ_m, $m=1,...,M$. Doing so, the joint distribution of the $\hat{\theta}_m$ is ignored, but a general, conservative solution (Nelson, 1989) for any type of correlation structure among the elementary $\hat{\theta}_m$ is provided.

$$1-P\left(\theta_m \in \left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right];\forall m=1,...,M\right) \leq \sum_{m=1}^{M} 1-P\left(\theta_m \in \left[\hat{\theta}_{mL};\hat{\theta}_{mU}\right]\right) \quad (4.1)$$

Figure 4.1 illustrates the Bonferroni-adjustment by a random sample of a bivariate Gaussian distribution in (4.2),

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad (4.2)$$

with $\rho = -0.5, 0, 0.5$. The solid lines are empirical (0.975)-quantiles q_1 and q_2 of the marginal distributions y_1 and y_2 obtained as outlined in Section 3.4. Interesting is

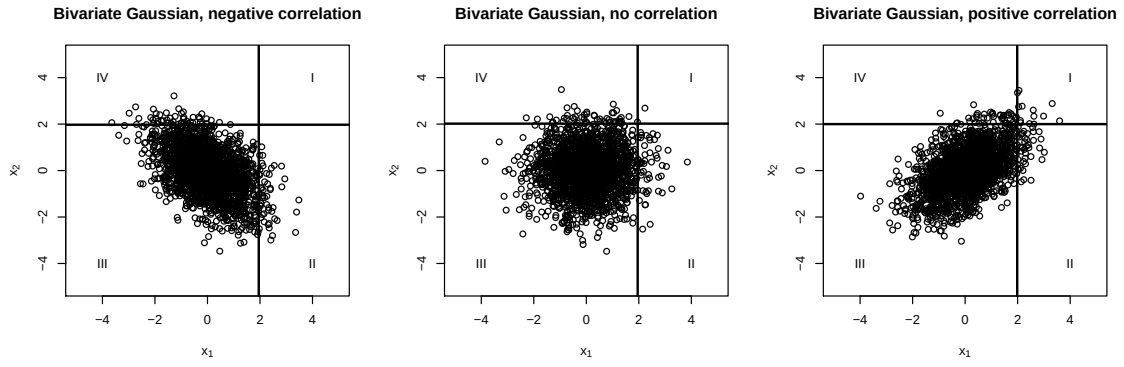

Figure 4.1: Scatter plots of 2000 realizations of bivariate Gaussian random variables with correlation $\rho = -0.5$ (left), $\rho = 0$ (center) and $\rho = 0.5$ and upper $(1 - 0.05/2)$-quantiles of the marginal distributions.

the proportion of the sample in which at least one of the variables x_1 and x_2 is larger than at least the corresponding quantile q_1, q_2 respectively (realizations in quadrants I, II and IV). This proportion is about 0.05 in the cases $\rho = -0.5$ and $\rho = 0$. With positive correlation $\rho = 0.5$, a certain proportion of the sample is larger than both, q_1 and q_2 (quadrant I). Hence, when q_1 and q_2 are chosen independently of each other, in order to cut off 2.5% of the marginal distributions, a part larger than 95% of the joint distribution is below both quantiles (quadrant III). The Bonferroni-adjustment leads to increasingly conservative intervals, when the probability mass in quadrant I increases, hence for increasingly positive correlations.

Sidak adjustment

The Sidak adjustment imposes the additional assumption that the elements of the estimator for $\boldsymbol{\theta}$ are mutually independent and preserves the simultaneous coverage probability whenever the true correlations are non-negative (Nelson, 1989). When mutually independent, the probabilities to contain the true parameters factorize for the elements of $\boldsymbol{\theta}$.

$$P\left(\theta_m \in \left[\hat{\theta}_{mL}; \hat{\theta}_{mU}\right]; \forall m = 1, ..., M\right) = \prod_{m=1}^{M} P\left(\theta_m \in \left[\hat{\theta}_{mL}; \hat{\theta}_{mU}\right]\right)$$

$$1 - \alpha = (1 - \alpha_m)^k$$

$$(1-\alpha)^{\frac{1}{k}} = 1-\alpha_m$$

Therefore, constructing confidence intervals with local confidence levels $(1-\alpha)^{1/k}$ for each θ_m, based on the marginal distribution lead to simultaneous confidence intervals with simultaneous confidence coefficient $(1-\alpha)$. The result is only slightly different from the Bonferroni adjustment and is conservative, when correlations are positive for the reasons illustrated in Figure 4.1. The Sidak procedure derived as above is suitable for one-sided inference; also two-sided versions have been proposed, see Nelson (1989).

4.3 Wald-type simultaneous confidence intervals based on approximation with the multivariate normal distribution

When it can be shown that for large N, that $\boldsymbol{T} \sim MVN(\mathbf{0}, \mathbf{R})$, with $\boldsymbol{T} = (T_1, ..., T_m, ..., T_M)'$, $T_m = \frac{\hat{\theta}_m - E(\theta_m)}{\sqrt{\hat{V}(\hat{\theta}_m)}}$, simultaneous Wald-type confidence intervals can be constructed by analytically solving Equation (4.3) for θ_{m0}:

$$\left| \frac{\hat{\theta}_m - \theta_{m0}}{\sqrt{\hat{V}\left(\hat{\theta}_m\right)}} \right| = z \tag{4.3}$$

Here, $z = z_{M,\mathbf{R},1-\alpha}^{two-sided}$ of an M-variate standard normal distribution which is chosen such that $P\left(|\mathbf{Z}| \leq z_{M,\mathbf{R},1-\alpha}^{two-sided}\right) = 1-\alpha$, where $\mathbf{Z}$ is an M-variate standard normal random vector with $(M \times M)$ correlation matrix $\mathbf{R}$. The confidence bounds are:

$$\left[\hat{\theta}_m^L; \hat{\theta}_m^U\right] = \left[\hat{\theta}_m \pm z\sqrt{\hat{V}\left(\hat{\theta}_m\right)}\right] \tag{4.4}$$

One-sided intervals can be obtained by using the critical value $z = z_{M,\mathbf{R},1-\alpha}^{one-sided}$ such that $P\left(\mathbf{Z} \leq z_{M,\mathbf{R},1-\alpha}^{one-sided}\right) = 1-\alpha$. Numerically, quantiles $z_{M,\mathbf{R},1-\alpha}$ can be obtained from the `R`-function `qmvnorm` in the add-on package `mvtnorm`, introduced by Hothorn et al. (2001). This adjustment takes the number of estimated parameters M, as well

as the correlation $\mathbf{R}$ among them into account.

For general problems, the correlation matrix $\mathbf{R}$ has elements $r_{mm'}$, $m \neq m'$, following from Equation (4.5):

$$r_{mm'} = \frac{C\hat{O}V\left(\hat{\theta}_m, \hat{\theta}_{m'}\right)}{\sqrt{\hat{V}\left(\hat{\theta}_m\right)\hat{V}\left(\hat{\theta}_{m'}\right)}} \tag{4.5}$$

When the normal approximation is sufficiently precise, such intervals achieve the objectives stated in Section 4.1. When the normal approximation is inappropriate due to small sample sizes, discrete, skewed, or kurtotic distributions of $\hat{\theta}_m$ the objectives in Section 4.1 are violated.

4.4 Simultaneous confidence intervals based on an empirical joint posterior distribution

While the construction of marginal confidence intervals from posterior distributions is standard in Bayesian application and software, the construction of simultaneous confidence sets is rarely considered. Suppose that MCMC has been used to draw samples from the joint posterior distribution of $\boldsymbol{\theta}$ given a sample y, the statistical model assumed for the creation of y given $\boldsymbol{\theta}$ and the prior assumptions on the joint distribution of $\boldsymbol{\theta}$. In the following, a sample of K values from the M-dimensional joint posterior will be denoted $\boldsymbol{\theta}^*$. $\boldsymbol{\theta}^*$ is written as a $(K \times M)$ matrix and will be called empirical joint posterior distribution of $\boldsymbol{\theta}$. Besag et al. (1995) describe the construction of simultaneous credible regions for $\boldsymbol{\theta}$ based on a sample of K realizations of an MCMC run, where the realizations are indexed by $k = 1, ..., K$. The region is intended to contain $100(1-\alpha)\%$ of the K realizations. For this purpose, the nearest integer to $K(1-\alpha)$ is chosen and denoted k^*. Denote θ_{km} be the kth value of the mth element of $\boldsymbol{\theta}^*$.

1. Order each of the M columns of $\boldsymbol{\theta}^*$ separately. Results are the order statistics $\theta_m^{[k]}$, and the ranks u_{km}, written in an $(K \times M)$ matrix $\mathbf{U}$. In that way, the empirical distributions of the elementary parameters $\theta_1, ..., \theta_m, ..., \theta_M$ are

brought into a comparable scale, the rank scale.

2. Calculate the minimum and maximum over each of the K rows of $\mathbf{U}$, $u_k^{(min)} = min\left(u_{k1}, ..., u_{km}, ..., u_{kM}\right)$, $u_k^{(max)} = max\left(u_{k1}, ..., u_{km}, ..., u_{kM}\right)$, and then calculate $u_k^{(maxmin)} = max\left(K + 1 - u_k^{(min)}, u_k^{(max)}\right)$, for each $k = 1, ..., K$. This step leads to the 'empirical distribution' of the maximum of the ranks after folding over the median. This is the basic step of making the credible intervals simultaneous.

3. The vector $\mathbf{u}^{(maxmin)} = \left(u_1^{(maxmin)}, ..., u_k^{(maxmin)}, ..., u_K^{(maxmin)}\right)$ is again ordered, leading to order statistics $u^{[k]}$ and the corresponding ranks $\mathbf{r}^{(k)}$. This allows for finding the (k^*)th quantile of the distribution of the maximum on the rank scale.

4. The 'critical value' or quantile is then $t^* = u^{[k^*]}$, taking the k^*th value from the ordered sample from the folded empirical distribution of the maximum. It has a similar function as the critical value $z_{M,\boldsymbol{R},1-\alpha}^{two-sided}$ in the parametric confidence intervals above, which also is the $(1-\alpha)$ quantile of the distribution of $max\left(|Z_1|, ..., |Z_m|, ..., |Z_M|\right)$.

5. Finally, the confidence limits are constructed for each elementary parameter θ_m by taking $\left[\theta_m^{[K+1-t^*]}; \theta_m^{[t^*]}\right]$, i.e. the $K + 1 - t^*$th and t^*th value from the ordered sample of the joint empirical distribution obtained for θ_m.

The derived region is two-sided for each parameter θ_m. Analogously, one-sided regions can be constructed. E.g., lower limits can be obtained in the following way:

1. Order $\boldsymbol{\theta}_*$ for each component M separately. Results are the order statistics $\theta_m^{[k]}$, and the ranks u_{km},

2. Calculate for each of the K rows, $k = 1, ..., K$, of $\mathbf{U}$ $u_k^{(min)} = K + 1 - min\left(u_{k1}, ..., u_{km}, ..., u_{kM}\right)$,

3. Order the vector $\mathbf{u}^{(min)} = \left(u_1^{(min)}, ..., u_k^{(min)}, ..., u_K^{(min)}\right)$, leading to order statistics $u^{[k]}$ and the corresponding ranks $\mathbf{r}^{(k)}$.

4. The 'critical value' or quantile is then $t^* = u^{[k^*]}$, taking the k^*th value from the

ordered sample from the empirical distribution of the minimum on the rank scale.

5. Then the one-sided lower M-dimensional credible set for $\boldsymbol{\theta}$ is constructed by taking $\left[\theta_m^{[K+1-t^*]}; \infty\right]$ for each $m = 1, ..., M$.

The subsets of $\boldsymbol{\theta}^*$ constructed in such a way have the following properties: For discrete joint empirical distributions, the SCI contain more than $(1-\alpha)K$ values of the K realizations. Hence, the confidence intervals derived in such a way are conservative in theory. This conservative performance increases with increasing dimension M (Besag et al., 1995). Confidence regions derived in such a way are non-parametric, in the sense that they do not depend on assumptions concerning the distribution of θ.

In initial investigations, Dilba (2006) showed the similarity of simultaneous credible intervals derived from MCMC with non-informative priors and those derived by frequentist analysis (Dilba, 2005; Dilba et al., 2006) in case studies assuming homoscedastic Gaussian response.

4.5 Parameters of interest in multiple treatment comparisons

Let $\boldsymbol{\mu}$, with elements μ_n, $n = 1, ..., N$ be the vector of parameters modeling the expectation of the observable response variable $\boldsymbol{y}$, with elements y_n, $n = 1, ..., N$ in dependence of a factor variable. Let $g\,()$ be a transforming function (link function); in most models considered in the following, the log-link is used, $g\,(\mu_n) = \log\,(\mu_n)$. Consider generalized linear models (McCullagh and Nelder, 1989) of the form:

$$g\,(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}, \tag{4.6}$$

where the variable defining the grouping of the observations $\boldsymbol{y}$ according to the I treatments is dummy coded in the $(N \times I)$ design matrix $\boldsymbol{X}$ with elements x_{ni} and $\boldsymbol{\beta}$ be an $(I \times 1)$ vector modeling the expectations of $\boldsymbol{y}$ in dependence of $\boldsymbol{X}$ on the

link. Further, let $\boldsymbol{\epsilon}$ denote an $(N \times 1)$ vector, containing the sums of all additional (random) effects which are not of primary interest.

Throughout this work, the design matrix $\boldsymbol{X}$ has the form of a cell means model, i.e., its elements have the value $x_{ni} = 1$ when observation y_n belonged to the ith treatment of the factor variable and $x_{ni} = 0$ otherwise. See (4.7) for an illustration.

$$\begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ \vdots \\ I \\ I \end{pmatrix} : \quad \boldsymbol{X} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{4.7}$$

With this parametrization, $\boldsymbol{\beta}$ models the means of levels $i = 1, ..., I$ on the scale of the link. However, when also numeric covariates are of interest, other choices are more appropriate, e.g. containing an intercept column, $I - 1$ columns for the differences to the intercept and columns containing the standardized values of the (standardized) covariates (Spiegelhalter et al., 2007).

As outlined in Section 1, interest is usually not directly in the elements of $\boldsymbol{\beta}$, β_i, but in parameters that express the dissimilarity among the I treatments, e.g. in terms of particular differences of elementary β_i or corresponding ratios on the original scale of expected values of $\boldsymbol{y}$, $\boldsymbol{\mu}$. For simplicity of the model definition and to allow a more general consideration of multiple comparison problems, the M-dimensional parameter vector of interest is defined in terms of $(M \times I)$ contrast matrices $\boldsymbol{C}$ with elements c_{mi}.

Canonical parameters for dissimilarity among groups

The canonical choice to express dissimilarity among groups in models assuming additivity on the link scale, is to define linear combinations, i.e., differences (of weighted arithmetic means) of the elements β_i, with weights c_{mi}. Then, the parameters of interest δ_m, $m = 1, ..., M$ are:

$$\delta_m = \sum_{i=1}^{I} c_{mi}\beta_i. \tag{4.8}$$

Note, that in case of generalized linear models with log-link, defining the parameter of interest as in Equation (4.8), i.e., as differences of expected values on the log-scale, results in ratios of (weighted geometric means of) expected values on the original scale after applying the exponential function to δ_m from Equation (4.8):

$$\rho_m = \exp\left(\delta_m\right). \tag{4.9}$$

If only pairwise comparisons are considered, using the ratio as parameter for dissimilarity among groups makes sense. In case that other contrasts than pairwise comparisons of treatments are defined in a $(M \times I)$ contrast matrices $\boldsymbol{C}$ with the constraints $\sum_{i=1}^{I} c_{mi} = 0, \forall m = 1, ..., M$ and $\sum_{i:c_{mi}>0} c_{mi} = 1, \forall m = 1, ..., M$, the parameter of interest resulting from Equations (4.8) and (4.9) can then be written as in Equation (4.10):

$$\rho_m = \frac{\prod_{i:c_{mi}\geq 0} \left[\exp\left(\beta_i\right)\right]^{c_{mi}}}{\prod_{i:c_{mi}<0} \left[\exp\left(\beta_i\right)\right]^{|c_{mi}|}}. \tag{4.10}$$

When the choice of the geometric mean to pool among several groups is not appropriate, one might also define ratios of (weighted) arithmetic means to compare pooled means of several groups.

Non-canonical parameters for dissimilarity among groups

In a Gaussian model with identity link, $g\left(\mu_n\right) = \mu_n$, interest may be in M ratios of (linear combinations of) the elements β_i. For that purpose, Dilba et al. (2006) define two $(M \times I)$ matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, with elements a_{mi} and b_{mi}, respectively. The ratios ρ_m, $m = 1, ..., M$ are then defined:

$$\rho_m = \frac{\sum_{i=1}^{I} a_{mi}\beta_i}{\sum_{i=1}^{I} b_{mi}\beta_i}. \tag{4.11}$$

However, the definition of ratios as in Equation (4.9) is considered as the canonical choice to describe dissimilarity among groups for strictly non-negative count data. Note, that in case of group-wise comparisons to control and all cases with $|c_{im}| = 1, \forall i = 1, ..., I \cap \forall m = 1, ..., M$ strictly fulfilled, the definition of ratios in Equation (4.9) does not differ from that in Equation (4.11).

Occasionally, one might be interested in the differences of (weighted arithmetic means of) expected values on the original scale also for models assuming additivity on the log scale:

$$\delta_m = \sum_{i=1}^{I} c_{im} \exp(\beta_i). \tag{4.12}$$

Comparisons to control

One of the most common experimental questions are multiple comparisons to a single control treatment, or comparisons of a single novel treatment to several control treatments. For comparisons to the first group, the $(M \times I)$, $M = I - 1$ contrasts matrix $\boldsymbol{C}$ with elements c_{mi} can be formally defined as:

$$c_{mi} = \begin{cases} -1 & \text{if i=1,} \\ 1 & \text{if i=m+1,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.13}$$

Such matrices yield differences to the control group ($i = 1$) when applied in Equations (4.8), (4.12) and ratios to control when applied in Equation (4.9). Exemplarily, for a setting with $I = 4$ treatments and $M = 3$ parameters of interest, the contrast matrix $\boldsymbol{C}$ is:

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \tag{4.14}$$

If interest is in ratios to control one might use the definition of Dilba et al. (2006):

$$a_{mi} = \begin{cases} 1 & \text{if i=m+1,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

$$b_{mi} = \begin{cases} 1 & \text{if i=1,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.16}$$

yielding

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.17}$$

and

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{4.18}$$

in the example with $I = 4$ treatments.

Further contrasts

The following choices of contrasts $\boldsymbol{C}$ are not of major importance in safety assessment. Nevertheless, they will be considered for a small number of settings in the simulation studies in order to expand the scope of this work to more general multiple comparison problems.

All pairwise comparisons are frequently of interest. In the context of multiple contrast tests the resulting parameters have a correlation matrix which is not of full rank (Bretz, 1999), hence are a problematic choice. For the example of $I = 4$ treatments, one yields

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \tag{4.19}$$

Further, trend tests can be expressed as multiple contrast tests (Bretz, 1999; Bretz and Hothorn, 2002). One example is the Williams-type contrasts for unbalanced

sample sizes introduced by Bretz (2006). With n_i denoting the sample size in the ith treatment, $i = 1, ..., I$, the contrast matrix $\boldsymbol{C}$ has the form:

$$\boldsymbol{C} = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 & 1 \\ -1 & 0 & \cdots & 0 & \frac{n_{I-1}}{n_{I-1}+n_I} & \frac{n_I}{n_{I-1}+n_I} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ -1 & \frac{n_2}{n_2+\cdots+n_I} & \cdots & \frac{n_{I-2}}{n_2+\cdots+n_I} & \frac{n_{I-1}}{n_2+\cdots+n_I} & \frac{n_I}{n_2+\cdots+n_I} \end{pmatrix} \tag{4.20}$$

Such contrasts differ from those mentioned before by defining differences of weighted arithmetic means when the parameter of interest is $\boldsymbol{\delta} = \boldsymbol{C\beta}$.

Joint posterior for the parameter of interest

Assume that MCMC is used to derive a sample of K values from the joint posterior distribution of a primary parameter vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_I)^t$, stored in a $(K \times I)$ matrix $\boldsymbol{\beta}^*$ with elements β^*_{ki}. A sample of K values from the joint posterior of the parameter of interest $\boldsymbol{\theta} = (\theta_1, ..., \theta_M)$ can be derived by applying

$$\theta^*_{km} = \sum_{i=1}^{I} c_{mi}\beta^*_{ki}, \tag{4.21}$$

when the parameter of interest can be defined as in (4.8) or (4.12), and

$$\theta^*_{km} = \frac{\sum_{i=1}^{I} a_{mi}\beta^*_{ki}}{\sum_{i=1}^{I} b_{mi}\beta^*_{ki}}, \tag{4.22}$$

when the parameter of interest can be defined as in (4.11), for all $k = 1, ..., K$, $m = 1, ..., M$. Likewise, the joint distribution for the parameters of interest in Equations (4.9) and (4.12) can be obtained.

4.6 SCI in a one-way model with Gaussian response

This section attempts to show that, in principle, MCMC can be used to construct simultaneous credible intervals with competitive frequentist performance, when a non-informative prior is assumed for the parameters of interest. However, the situations of primary interest (Sections 6 and 7) are characterized by discrete distributions in complex models. These situations also involve the problem of jointly

estimating the expectation and variance, in situations where these moments jointly depend on several parameters. Hence, problems of estimation which might be difficult irrespective of the used method. In complex hierarchical models, technically inappropriate choices of parameters in the update process of MCMC may cause convergence problems and autocorrelations might lead to inappropriate approximations of joint empirical distributions. Hence, the proof of concept is tried for the well behaved problem of multiple comparisons in the one-way layout with homoscedastic Gaussian response. As an example, Dunnett-type comparisons to control are considered. Both the difference and the ratio are taken as measures of dissimilarity, for which frequentist solutions have been published long ago (Dunnett, 1955) and recently (Dilba et al., 2006), respectively.

4.6.1 Model

Assume the following model:

$$\begin{aligned} Y_n &= \mu_n + \epsilon_n \\ \mu_n &= \textstyle\sum_{i=1}^{I} x_{ni}\beta_i \\ \epsilon_n &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{4.23}$$

Non-informative priors are assumed for both $\boldsymbol{\beta}$ and σ: $\beta_i \sim N\left(0, 1000\right)$, and $\tau = 1/\sigma^2 \sim Gamma\left(a = 0.001, b = 1000\right)$. The `BUGS` code defining this model and the technical details of the MCMC updates can be found in Section 4.6.4.

4.6.2 An example

As a comparison on the basis of realized confidence intervals, the *Angina pectoris* data set of Westfall et al. (1999, p.164) is considered in detail. The data set (Summary statistics in Table 4.1) is a dose response study of a drug to treat *Angina pectoris*, with an untreated control (Dose 0) and four doses (Dose 1,...,4). The response variable is the difference post treatment - pre treatment of the duration of pain-free walking, i.e, large values indicate positive drug effects on the patients.

Table 4.1: Summary statistics of the *Angina pectoris* data set: sample size if the ith treatment, n_i, arithmetic mean of the ith sample, $\hat{y}_i$, standard deviation of the ith sample $\hat{\sigma}_i$, minimal $\min(y_i)$ and maximal $\max(y_i)$ value.

Dose	n_i	$\hat{y}_i$	$\hat{\sigma}_i$	$\min(y_i)$	$\max(y_i)$
0	10	14.1	3.1	10.35	19.06
1	10	16.2	4.0	9.63	23.78
2	10	17.5	2.7	13.51	22.86
3	10	19.1	3.2	13.03	23.38
4	10	24.6	4.1	18.25	32.32

Assuming model (4.23) for the data leads to a primary parameter vector with elements β_i, i=1,...,5. Interest is in the ratios and differences to control, as defined in Equations (4.11) and (4.8), respectively. Applying the `BUGS` code in Section 4.6.4 (with one additional column in the design matrix), running 4000 updates, discarding the first 2000 updates and every second update in the remaining part yields a joint posterior distribution $\boldsymbol{\beta}^*$ for which all bivariate projections are shown in Figure 4.6.2. The independence assumed for the five samples is reflected in the apparently uncorrelated joint distributions of the β_i. Applying Equation (4.21) with a (4×5) Dunnett-type contrast matrix $\boldsymbol{C}$ and Equation (4.22), yields joint posterior distributions for the parameters of interest as shown in Figures 4.6.2 and 4.6.2, respectively. The obvious correlation of the parameters of interest in Figure 4.6.2 reflects the theoretically expected correlation $r_{mm'} = 0.5$ for $m \neq m'$ in the case of differences to control (Dunnett, 1955). As can be expected from the results of Dilba et al. (2006), the correlations among the ratios to control is slightly higher and differs slightly for the different comparisons, since the observed ratios are increasingly greater than 1 with increasing dose. Finally, applying the method outlined in Section 4.4 to the joint posterior distributions yields the simultaneous confidence intervals shown in Table 4.2. This table also shows the bounds obtainable by directly using Dunnett (1955) and Dilba et al. (2006) based on sample estimates for β_i and σ. The bounds obtained by MCMC appear practically equivalent to those obtained by the standard

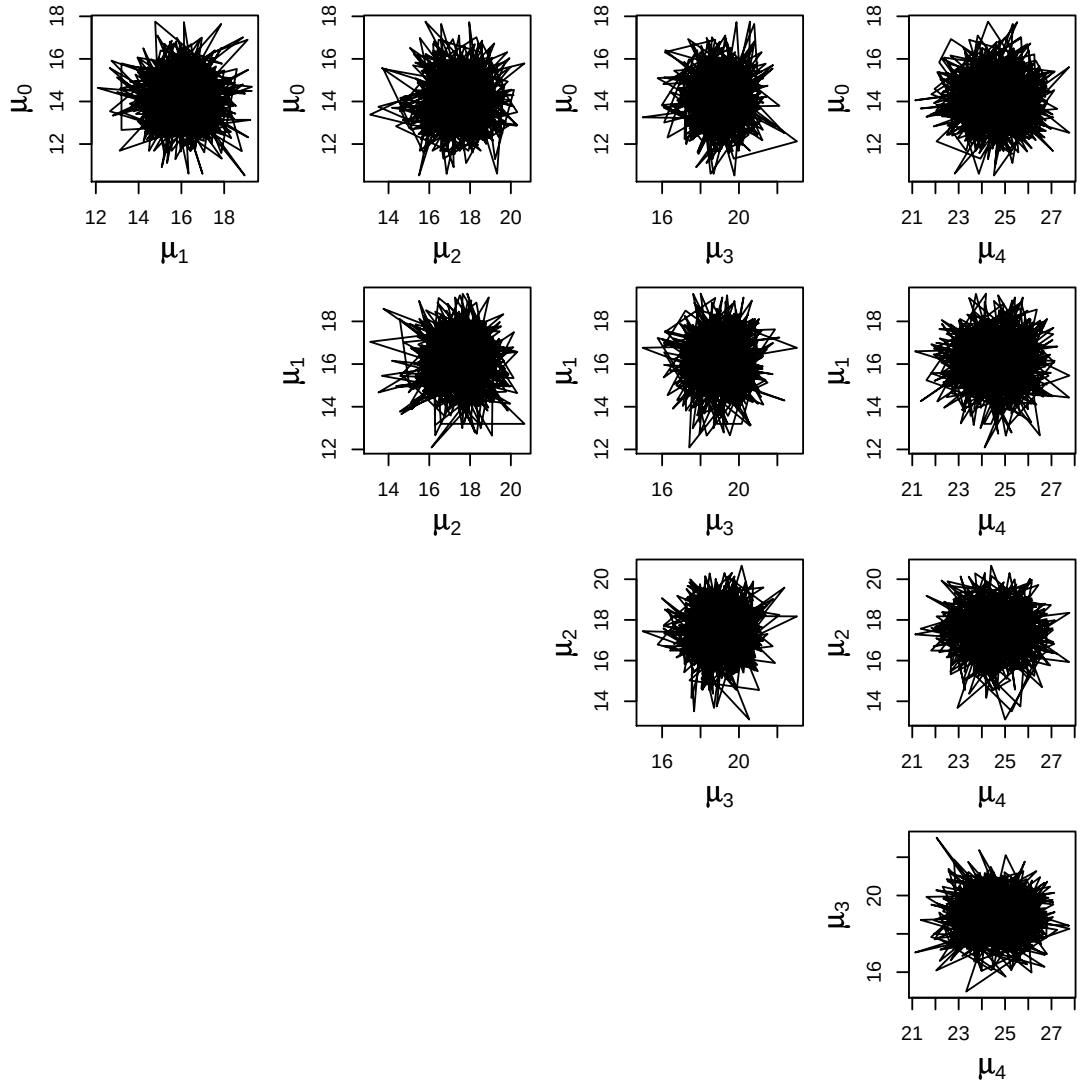

Figure 4.2: Bivariate projections of the MCMC sample of the joint distribution of the $I = 5$ means of the treatment groups. Lines join consecutive draws in the MCMC sample.

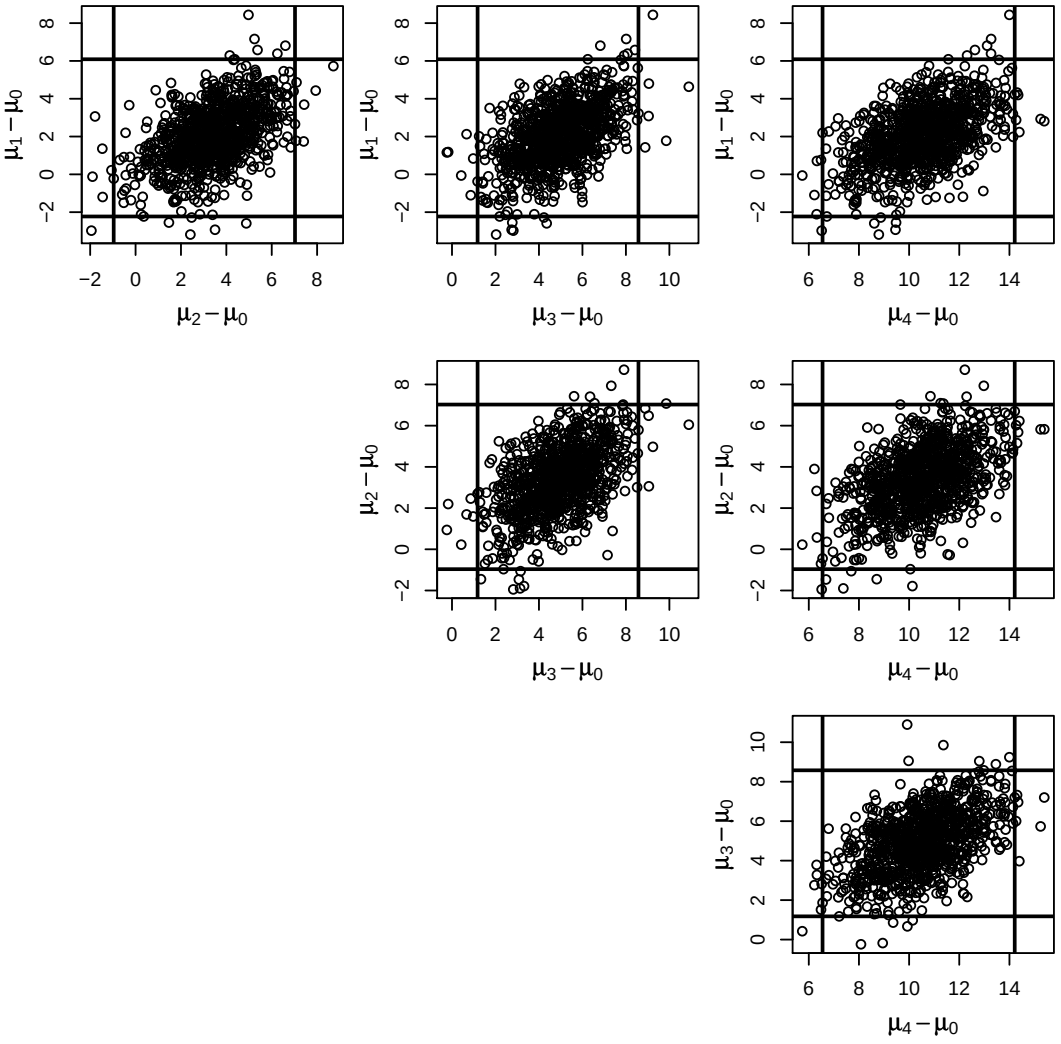

Figure 4.3: Scatter plots of bivariate projections of the MCMC sample of the joint distribution of the $M = 4$ differences of means of the dose groups to the mean of the untreated groups. The solid lines indicate the limits of 0.95 confidence sets derived by MCMC.

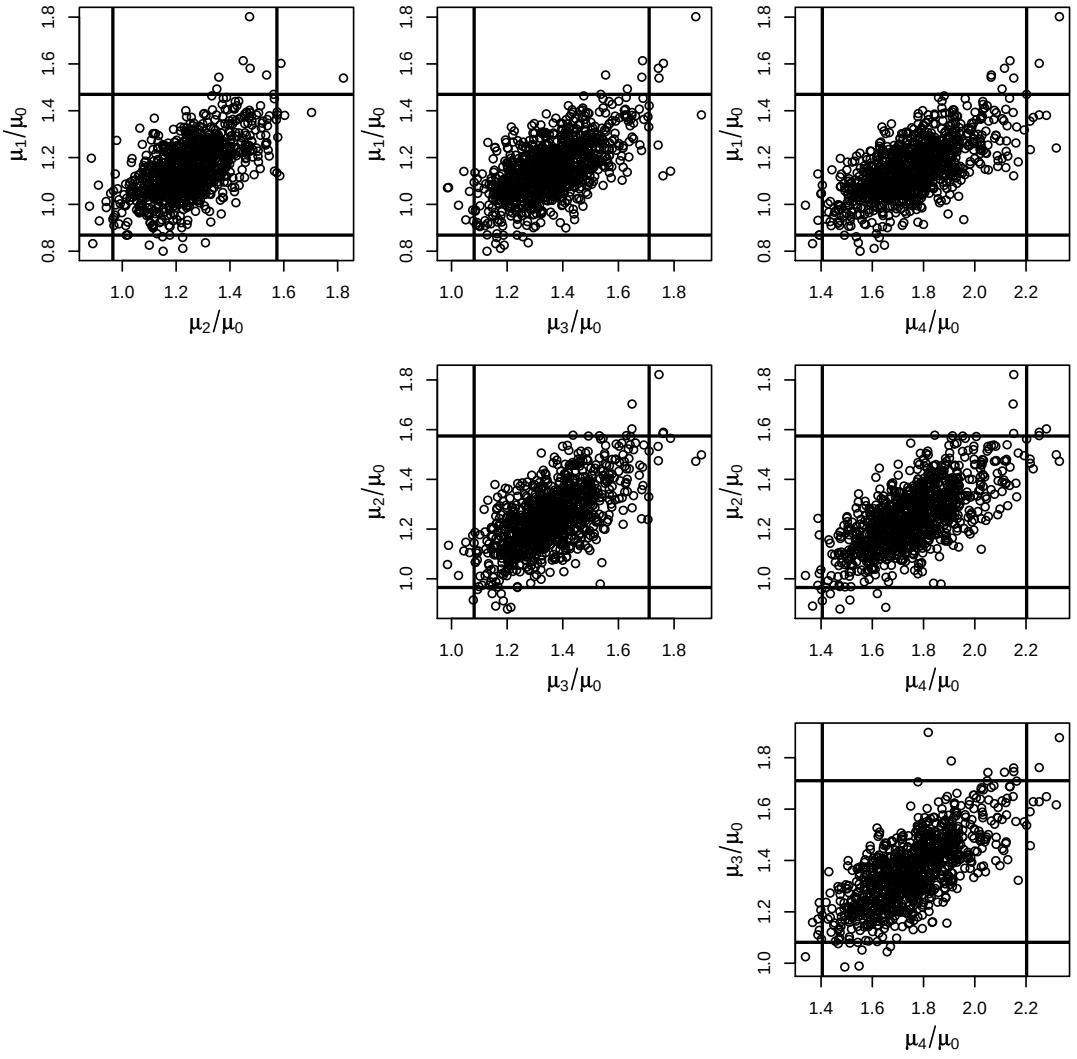

Figure 4.4: Scatter plots of bivariate projections of the MCMC sample of the joint distribution of the $M = 4$ ratios of means of the dose groups to the mean of the untreated groups. The solid lines indicate the limits of 0.95 confidence sets derived by MCMC.

methods.

Table 4.2: Two-sided 0.95 confidence intervals for the $M = 4$ comparisons to control based on the ratios and differences of means. The first column names the parameter, column two gives the sample estimate, columns three and four show the lower and upper bounds of the simultaneous (0.95) confidence intervals according to the frequentist solutions of Dilba et al. (2006) and Dunnett (1955), columns five and six show lower and upper bounds of simultaneous (0.95) confidence intervals derived from MCMC sampling.

		Frequentist		MCMC	
Parameter	Estimate	Lower	Upper	Lower	Upper
μ_1/μ_0	1.149	0.890	1.496	0.869	1.470
μ_2/μ_0	1.241	0.972	1.606	0.965	1.574
μ_3/μ_0	1.354	1.071	1.742	1.081	1.710
μ_4/μ_0	1.745	1.411	2.213	1.405	2.202
$\mu_1 - \mu_0$	2.095	−1.825	6.015	−2.227	6.088
$\mu_2 - \mu_0$	3.397	−0.523	7.317	−0.964	7.025
$\mu_3 - \mu_0$	4.995	1.075	8.915	1.177	8.577
$\mu_4 - \mu_0$	10.499	6.579	14.419	6.549	14.211

4.6.3 Simulation study: Summary of results

Table 4.3 in Section 4.6.5 shows simulation results for 15 settings with balanced sample sizes $n_i = 5, 10, 100$ for nominal two-sided 0.95 SCI and their lower and upper 0.975 bounds, respectively. Simultaneous confidence intervals were computed for $\boldsymbol{\delta}$ and $\boldsymbol{\rho}$ as defined in (4.13) and (4.15, 4.16), respectively. The coverage probability is close to the nominal level for small sample sizes $n_i = 5$ for two-sided as well as one-sided consideration. For larger sample sizes, the SCI become slightly liberal in tendency.

Here it is found that simultaneous credible intervals from empirical joint distributions obtained from MCMC may have acceptable frequentist coverage probability, when non-informative priors are imposed on the parameters of interest.

4.6.4 BUGS code and update parameters

Consider a one way layout with homoscedastic Gaussian response $I = 4$ groups and non-informative priors on the means and the variance parameter as defined in (4.23). Y is the $(N \times 1)$ matrix of observations, the $(N \times 4)$ design matrix X and the corresponding (4×1) parameter vector β are assumed to be parameterized as a cellmeans model. Note that the normal distribution is parameterized with expectation `mu` and precision parameter `tau`, the reciprocal of the variance.

```
model
{
for(n in 1:N){
X[n,1] <- X1[n]
X[n,2] <- X2[n]
X[n,3] <- X3[n]
X[n,4] <- X4[n]
Y[n] ~ dnorm(mu[n], tau)
mu[n] <- inprod(X[n,], beta[])
```

```
}
for(p in 1:P){
 beta[p] ~ dnorm(0,0.0001)
 muvec[p]<-beta[p]
}
tau ~ dgamma(0.001, 0.001)
sigma<-1/sqrt(tau)
}
```

A similar model can be found in the `OpenBUGS` example manual (Spiegelhalter et al., 2007). For updating the model, the vectors `Y`, `X1`, `X2`, `X3`, `X4` and the integers `N` and `P` need to be provided along with initial values for the vector `beta` and the real `tau`. The simulation has been run with S=1000 Monte Carlo draws from the assumed model for each parameter setting. For each random sample, SCI were computed based on an MCMC run with 2000 updates, first 1000 updates discarded and no thinning, resulting in a sample of $K = 1000$ values from the joint posterior.

4.6.5 Detailed results

Table 4.3: Simultaneous coverage probability of two-sided nominal 0.95 SCI derived from MCMC with $\sigma = 1$, for Dunnett-type contrasts for the ratio ρ and difference δ of means of Gaussian samples. Simulations based on $K = 1000$ and S=1000.

					ρ			δ		
μ_1	μ_2	μ_3	μ_4	n_i	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50.0	50.0	50.0	50.0	5	0.952	0.977	0.974	0.952	0.977	0.974
50.0	62.5	50.0	62.5	5	0.943	0.973	0.970	0.940	0.969	0.971
50.0	40.0	50.0	40.0	5	0.944	0.970	0.974	0.941	0.969	0.972
50.0	50.0	50.0	250.0	5	0.955	0.975	0.980	0.951	0.978	0.973
50.0	50.0	50.0	5.0	5	0.949	0.978	0.971	0.942	0.968	0.974
50.0	50.0	50.0	50.0	10	0.934	0.956	0.978	0.934	0.956	0.978
50.0	62.5	50.0	62.5	10	0.939	0.963	0.976	0.942	0.967	0.975
50.0	40.0	50.0	40.0	10	0.939	0.966	0.973	0.940	0.967	0.973
50.0	50.0	50.0	250.0	10	0.961	0.978	0.983	0.950	0.976	0.974
50.0	50.0	50.0	5.0	10	0.955	0.975	0.980	0.951	0.974	0.977
50.0	50.0	50.0	50.0	100	0.946	0.963	0.983	0.946	0.963	0.983
50.0	62.5	50.0	62.5	100	0.937	0.964	0.973	0.937	0.962	0.975
50.0	40.0	50.0	40.0	100	0.936	0.957	0.978	0.941	0.962	0.979
50.0	50.0	50.0	250.0	100	0.938	0.963	0.975	0.931	0.953	0.978
50.0	50.0	50.0	5.0	100	0.935	0.965	0.970	0.944	0.966	0.978

Chapter 5

Confidence intervals for ratios of means of negative binomials in the one-way layout

In this Chapter, methods formally described in Chapter 3 are applied to the problem of constructing local and simultaneous confidence intervals for contrasts of group-wise means, where the response variable y is assumed to follow a negative binomial distribution.

5.1 Statistical model

Assume a simple one-way layout with a response vector $\boldsymbol{y}$, containing N mutually independent observations y_n, which are classified by a single variable with I levels, $i = 1, ..., I$. The classifying variable is dummy coded $(0, 1)$ in an $(N \times I)$ design matrix $\boldsymbol{X}$, and $\boldsymbol{\beta}$ is a $(I \times 1)$ parameter vector, estimating the group means on the log-scale. For simplicity, it is assumed that $\boldsymbol{X}$ is defined as outlined in Section 4.5, Equation (4.7). It is assumed that the variability of $\boldsymbol{y}$ can be entirely described by the unknown expectation $\boldsymbol{\mu}$ and the assumption of the negative binomial distribution with unknown but common dispersion parameter τ. The model can be formally

defined by Equation (5.1):

$$\begin{aligned} Y_n &\sim NB(\mu_n, \tau) \\ \log(\mu_n) &= \eta_n \\ \eta_n &= \sum_{i=1}^{I} x_{ni}\beta_i \end{aligned} \tag{5.1}$$

In the simplest case, consider the model in Equation 5.1 with $I = 2$, where interest is in confidence intervals for the ratio of the means, $\theta = \mu_2/\mu_1$.

5.1.1 Model fit and estimation of the dispersion parameter

In the frequentist setting, the above model can be fitted using maximum likelihood methods as described in Venables and Ripley (2002) or the algorithm proposed by Rigby and Stasinopoulos (2005). However, the problem involves joint estimation of mean and dispersion parameters, where the variance of $\boldsymbol{Y}$ depend on the parameters μ and τ in Equation (A.7) or on a and b in Equation (A.6). Saha and Paul (2005) investigated estimation procedures for the dispersion parameters. They come to the conclusion that different previously proposed methods lead to negatively biased estimators. Consequently, they propose a bias-corrected estimator which, however, still shows negative bias for relevant parameter settings.

5.1.2 Inference for negative binomial parameters

Although the negative binomial distribution has been used since long to model overdispersed count data (Anscombe, 1949, 1950; Bliss and Fisher, 1953), the validity of inferential methods following the model fit in general and the coverage probability of local and simultaneous confidence intervals for dissimilarity of μ_i in particular has rarely been considered. Lawless (1987) recommends the normal approximation for estimating μ for samples as small as $n = 25, 30$. Simulation studies show that the distribution of sample estimates for μ_i in two-parameter problems has heavier tails than a standard normal distribution. Breslow (1990) considers tests for main

effects (tests for the significance of added variables) in log-linear models assuming negative binomial response among others; he shows that tests are conservative when overdispersion is negligible, but become liberal when there is marked overdispersion. Campbell et al. (1999) consider the size of F-tests and χ^2-tests for two-factorial additive models assuming the negative binomial distribution; they again find that large overdispersion leads to markedly liberal tests while for settings close to the Poisson distribution, the considered methods become conservative. Gerhard and Schaarschmidt (2007) show simulation results for the coverage probability of Wald-type intervals for the ratio of two means, with results corresponding to those of Breslow (1990) and Campbell et al. (1999). Recently, Hothorn et al. (2008) consider asymptotic inference for multiple treatment comparisons in generalized linear models; however, small sample properties of the described simultaneous Wald-type confidence intervals are not considered.

5.2 Wald-type confidence intervals

Fitting the model in Equation (5.1) yields estimates for the expectation on the log-scale $\hat{\boldsymbol{\beta}}$ with elements $\hat{\beta}_1, \hat{\beta}_2$ and the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ with the diagonal elements $\hat{\sigma}_1, \hat{\sigma}_2$ (McCulloch and Searle, 2001). Assuming completely randomized designs and a parametrization as outlined in (4.7), the off-diagonal elements of $\boldsymbol{\Sigma}$ are 0 and marginal $(1 - 2\alpha)$ Wald-type confidence intervals can be constructed using Equation (5.2), (McCulloch and Searle, 2001):

$$\left[\hat{\theta}^l; \hat{\theta}^u\right] = \exp\left(\hat{\beta}_2 - \hat{\beta}_1 \pm z_{1-\alpha}\sqrt{\hat{\sigma}_2^2 + \hat{\sigma}_1^2}\right), \tag{5.2}$$

where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution. Gerhard and Schaarschmidt (2007) showed by simulation studies that Wald intervals for the considered model may be liberal when there is clear overdispersion and sample sizes are small (5-10 observations per group). For other parameter settings with small sample sizes, the intervals may be very conservative, namely covering the true parameter with probability close to 1.

5.3 Confidence intervals based on MCMC

One can translate the statistical model in (5.1) for the case $I = 2$ into the `BUGS` model in Section 6.3.2 with non-informative priors imposed on all parameters. An empirical distribution for θ is obtained by recording the K values of $\exp(\beta_2 - \beta_1)$ of an MCMC run with K updates. Confidence intervals can then be constructed as described in Section 3.4.

5.4 Performance for observing only zeros

The result of observing only zeros in at least one sample is a common problem in overdispersed count data (e.g. Prescher, 2005, personal communication; Rauschen et al., 2008). For small samples, large overdispersion and small mean abundance, the case may occur that no individual is counted in at least one of the samples (see Table 5.1). Since that event is a reasonable outcome for a negative binomial random variable, and is not contradictory to the negative binomial assumption, it can be required that also in this particular case, confidence intervals for $\rho = \mu_i/\mu_{i'}$ are obtained that have the properties outlined in Section 3.1.

Table 5.1: Probability to observe only zeros in a sample of N independent draws from a negative binomial distribution with parameters μ and τ as defined in Equation (A.7) in the Appendix.

	Expectation μ						
	0.10	0.50	0.80	1.00	1.25	2.00	10.00
$N = 5, \tau = 1$	0.621	0.132	0.053	0.031	0.017	0.004	0.000
$N = 10, \tau = 1$	0.386	0.017	0.003	0.001	0.000	0.000	0.000
$N = 5, \tau = 10$	0.608	0.087	0.021	0.009	0.003	0.000	0.000
$N = 10, \tau = 10$	0.370	0.008	0.000	0.000	0.000	0.000	0.000

Consider the case of two independent samples, $\boldsymbol{y}_1$ and $\boldsymbol{y}_0$, each consisting of eight observations (Table 5.2). Consider now the problem of estimating two-sided nominal

Table 5.2: Hypothetical example with only 0 counts in one sample.

$\boldsymbol{y}_0 = (y_1, ..., y_8)'$	0	0	0	0	0	0	0	0
$\boldsymbol{y}_1 = (y_9, ..., y_{16})'$	0	0	0	2	0	0	0	1

0.9 confidence intervals for the ratio of mean abundance in sample y_1 relative to the mean abundance in sample y_0, μ_0/μ_1. Table 5.3 shows the resulting lower and upper 0.95 confidence limits of four methods: First, the Wald-type interval according to (5.2), based on estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ obtained from the algorithm of Rigby and Stasinopoulos (2005). Second, an MCMC derived interval using the `BUGS` code in Section 5.6.1 with a non-informative prior on β_i,

$$\beta_i \sim N\left(0, 1000\right). \tag{5.3}$$

Third, an MCMC derived interval using the `BUGS` code in Section 5.6.1 with a weakly informative prior on β_i,

$$\beta_i \sim N\left(0, 10\right). \tag{5.4}$$

Both MCMC derived intervals rely on a sample of $K = 1000$ values from the posterior, obtained from a single chain of 4000 values, with 2000 values discarded in the beginning and one out of two values discarded in the remaining part of the chain. Finally, an interval based on a likelihood profile is shown, using R code (Gerhard, 2008, personal communication; Venables and Ripley, 2002) which makes the calculation of likelihood profiles feasible for extreme events by restricting the parameter space on the log scale to $[-25, 25]$, being implemented in the package `pairwiseCI` (Schaarschmidt, 2008).

Obviously, the Wald-type interval covers practically the whole parameter space $[0, \infty]$, resulting from the simple problem, that the estimate for β_0 goes to $-\infty$ and the variance estimate at the position of the point estimate is (numerically close to) ∞. Using MCMC with a non-informative prior in Equation (5.3) results in a rather narrow confidence interval $[0; 0.054]$, implying a rather clear dissimilarity based on the weak empirical basis provided by the data. Using MCMC with the weakly informative prior in Equation (5.4) results in an interval with both bounds

Table 5.3: Lower and upper 0.95 confidence limits for the ratio of mean abundances μ_0/μ_1, based on the samples in Table 5.2, according to four different methods.

CI method	Estimate	Lower 0.95 limit	Upper 0.95 limit
Wald-type	2.2e−05	2.2e−58	2.3e+48
MCMC, non-informative prior	2.2e−09	3.3e−26	5.4e−02
MCMC, weakly informative prior	7.2e−02	2.4e−03	9.6e−01
Likelihood profile interval	1.5e−09	0.0e+00	6.8e−01

closer to 1 than the previous, $[0.002; 0.96]$. Here, the lower bound is mainly governed by the prior assumption on the means. The likelihood profile method yields $[0; 0.68]$, with the upper bound closest to the MCMC solution based on the weakly informative prior.

For the parameter μ_1/μ_0, similar inconsistencies between the methods are observed (Table 5.4). The Wald-type interval again practically covers the parameter space and fails to reflect that the evidence for $\mu_1 > \mu_0$ is slightly larger than the evidence for $\mu_1 < \mu_0$ after observing the sample.

Table 5.4: Lower and upper 0.95 confidence limits for the ratio of mean abundances μ_1/μ_0, based on the samples in Table 5.2, according to four different methods.

CI method	Estimate	Lower 0.95 limit	Upper 0.95 limit
Wald-type	4.5e+04	4.4e−49	4.6e+57
MCMC, non-informative prior	7.9e+08	2.2e+01	3.2e+25
MCMC, weakly informative prior	1.2e+01	1.3e+00	4.5e+02
Likelihood profile interval	6.7e+08	1.5e+00	∞

From the above considerations of a case where one sample consists only of 0-counts it is obvious that Wald-type confidence intervals yield inappropriate results for this important special case. Bounds close to $[0, \infty]$ appear whenever one sample contains only zeros, even when the other sample contains many non-zero counts, and therefore

provides substantial evidence for a dissimilarity in means. In a frequentist simulation of the coverage probability, this problem will show up as conservative performance (coverage probability larger than the prespecified level $(1 - \alpha)$), when settings with small sample size and mean abundance and large overdispersion are considered (see Gerhard and Schaarschmidt, 2007). In the two-sample case, point-wise inversion of a likelihood ratio test can be a remedy to this problem. MCMC derived intervals based on non-informative priors may yield too narrow confidence intervals.

5.5 Frequentist performance of MCMC derived confidence intervals: Simulation study

In simulation studies, the frequentist performance of MCMC derived marginal confidence intervals was assessed. Several models, slightly differing in the definition of prior distributions, were considered.

Model (5.1) can be written as the `BUGS` model in Section 5.6.1. In the simulation study with details shown in Section 5.6, for fixed values of $\beta_i, i = 1, 2$ and τ, the coverage probability of confidence intervals for $\rho = \exp(\beta_2 - \beta_1)$ and the non-canonical parameter $\delta = \exp(\beta_2) - \exp(\beta_1)$ was assessed for settings with balanced group-wise sample size 10, 20, 50 (i.e., $N = 20, 40, 100$ equally distributed to the $I = 2$ treatments). In the simulations, two-sided nominal 0.9 confidence intervals were calculated and frequentist coverage probabilities were estimated based on 10000 or 1000 replications of the random experiment.

Along with assessing the frequentist performance of the MCMC derived confidence intervals for different parameter values, technical parameters (number of updates, number of values discarded) of the MCMC runs were varied in order to assess whether the chosen technical parameters have influence on the observed performance.

5.5.1 Effect of mean abundance and sample size

Tables 5.6, 5.7 and 5.8 show that, with noticeable overdispersion $\tau = 1$, MCMC derived confidence interval for the ratio of means have a coverage probability decreasing with decreasing mean abundance and decreasing sample size. About nominal coverage probability is achieved when the mean abundance is large or intermediate $\mu_1 = 50, 10$ in combination with large group-wise sample sizes of 50. For intermediate group-wise sample sizes 20 the coverage probability may still be acceptable with large mean abundances but can be as small as 0.8 for nominal two-sided 0.9 intervals, when the mean abundance is about $\mu_i = 1$. For the often realistic sample size of 10 (Table 5.6), coverage probabilities are slightly below the nominal level (0.88-0.89) for

large abundances but clearly below the nominal level (0.84-0.88) when abundance is only intermediate and severely liberal (0.6-0.85) for abundances close to 1.

5.5.2 Upper and lower bounds

The coverage probabilities of lower and upper bounds separately are about equal, except in cases where group-wise sample sizes are small and the ratios extreme ($\rho = 0.1, 10$). Hence two-sided 0.9 confidence intervals for the ratio of means might be used for a proof of safety with level approximately 0.05, provided that sample sizes are at least 20 per group, mean abundance is not very low and the dissimilarity between the groups is not extreme.

5.5.3 Results for the difference of means

Considering the non-canonical difference of means yields similar overall conclusions concerning the validity of the confidence intervals, but slightly different results in the actual violations of the nominal levels. Also here, the coverage probabilities of lower and upper bounds separately are close to 0.95 for nominal two-sided 0.9 intervals, as long as the group-wise sample size is at least 20, the abundance is not too small (at least about 10).

5.5.4 Moderate overdispersion

Comparing the above results to situations with relatively moderate overdispersion with $\tau = 10$ (Tables 5.10 and 5.9 for group-wise sample sizes 20 and 10, respectively) the intervals are less liberal when abundances are low, but are about as liberal or more liberal than in a situation with high overdispersion.

5.5.5 Effect of using a gamma prior on τ

Table 5.5 contains results for defining the `BUGS` code with a non-informative uniform prior for the precision parameter $\tau \sim unif\,(0, 1000)$ while Table 5.11 shows results for the same parameter settings and model parameters using a non-informative gamma prior $\tau \sim gamma\,(0.001, 1000)$ for the precision parameter. For the most relevant setting of group-wise sample sizes 10, the confidence intervals based on a model with gamma prior are uniformly closer to the nominal level. Although the use of the gamma prior does not avoid the extremely liberal performance when mean abundance is very low ($\mu_2 = 0.1$) it has coverage probabilities 0.86-0.89 for situations with low abundance where the model using a uniform prior results in coverage probabilities 0.76-0.85. Based on the few simulated settings, one may carefully recommend to use a gamma prior instead a uniform prior.

5.5.6 Effects of number of updates

Comparing Table 5.5 with the results in Table 5.6 shows that a sufficient number of updates was used in the MCMC runs underlying the results discussed above. Decreasing the number of updates to 2000 and basing the confidence intervals on a sample of $K = 1000$ values from the posterior results in about equal estimates for the coverage probability. Additionally (results not shown), considering Gewekes (Gelman et al., 2004; Geweke, 1992) tests for convergence on the simulation results showed that the null-hypotheses is rejected less often than can be expected under the null-hypotheses of a stationary chain. When those simulations were omitted for which Gewekes test rejected the null-hypotheses for at least one parameter in the model or all parameters in the model, the resulting estimates for the coverage probability did not differ markedly from those shown in the Tables in Section 5.6.

5.5.7 Summary

MCMC-based confidence intervals for the ratio of two means assuming a negative binomial response show a liberal performance for the considered settings. Especially for very small mean abundance, the probability to cover the true parameter is substantially lower than the nominal level. When applied in a proof of safety, such confidence intervals will lead too often to a conclusion for safety when indeed the null-hypothesis of hazardousness is true.

5.6 A uniform prior for the dispersion parameter

5.6.1 BUGS code and update parameters

The following BUGS code represents the one-way model with negative binomial response of Equation (5.1), with $I = 2$ groups. The parametrization of the negative binomial and normal distribution as used in the following implementations in BUGS is explained in Appendix A.

```
model
{
for(n in 1:N)
{
X[n,1] <- X1[n]
X[n,2] <- X2[n]
Y[n] ~ dnegbin(pi[n], r)
pi[n] <- r/(r+mu[n])
mu[n] <- exp(eta[n])
eta[n] <- inprod(X[n,], beta[])
}
for(p in 1:P)
{
```

```
beta[p] ~ dnorm(0, 0.001)
muvec[p]<-exp(beta[p])
}
r ~ dunif(0,1000)
}
```

For running the model, the vectors `Y`, `X1`, `X2`, and the integers `N` and `P` have to be provided along with initial values for the vector `beta` with two elements and the real `r`.

5.6.2 Detailed results for $K = 1000$

In the following simulations, a chain of 2000 updates was run, with the first 1000 values discarded and no thinning, resulting in a sample of $K = 1000$ values from the posterior distribution. The estimates of the coverage probability are based on S=10000 random draws of $\boldsymbol{y}$ from the model in Equation (5.1) with $I = 2$, $\tau = 1$ and further parameters as indicated in the headers of the tables. Table 5.5 shows coverage probabilities for high, intermediate and low abundances and sample size $n_i = 10$.

Table 5.5: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio ρ and difference δ of means, assuming the negative binomial model in Equation (5.1), based on $K = 1000$ and $S = 10000$.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	10	0.1	−45.0	0.863	0.933	0.930	0.860	0.924	0.935
50	10	0.5	−25.0	0.873	0.937	0.936	0.873	0.928	0.945
50	10	0.8	−10.0	0.873	0.938	0.935	0.873	0.935	0.938
50	10	1	0.0	0.875	0.939	0.936	0.876	0.940	0.936
50	10	1.2	10.0	0.875	0.938	0.938	0.878	0.942	0.936
50	10	2	50.0	0.875	0.937	0.938	0.884	0.949	0.934
50	10	5	200.0	0.875	0.935	0.940	0.885	0.947	0.938
10	10	0.1	−9.0	0.850	0.941	0.908	0.836	0.904	0.932
10	10	0.5	−5.0	0.866	0.938	0.928	0.868	0.926	0.942
10	10	0.8	−2.0	0.868	0.937	0.931	0.870	0.934	0.936
10	10	1	0.0	0.867	0.936	0.931	0.867	0.936	0.931
10	10	1.2	2.0	0.872	0.939	0.933	0.871	0.941	0.930
10	10	2	10.0	0.872	0.939	0.933	0.873	0.948	0.925
10	10	5	40.0	0.869	0.937	0.932	0.875	0.947	0.928
10	10	10	90.0	0.878	0.939	0.940	0.883	0.943	0.940
1	10	0.1	−0.9	0.570	0.950	0.620	0.743	0.836	0.907
1	10	0.5	−0.5	0.765	0.892	0.873	0.768	0.869	0.899
1	10	0.8	−0.2	0.761	0.881	0.881	0.758	0.871	0.886
1	10	1.0	0.0	0.762	0.881	0.881	0.762	0.881	0.881
1	10	1.2	0.2	0.758	0.878	0.880	0.756	0.882	0.874
1	10	2.0	1.0	0.765	0.882	0.883	0.743	0.900	0.843
1	10	5.0	4.0	0.819	0.902	0.916	0.775	0.930	0.845
1	10	10.0	9.0	0.853	0.915	0.938	0.834	0.935	0.900

5.6.3 Detailed results for $K = 5000$

The following Tables 5.6, 5.7 and 5.8 show coverage probabilities, based on a chain of 6000 updates, with the first 1000 values discarded and no thinning, resulting in a sample of $K = 5000$ values from the posterior distribution. Coverage probabilities are based on $S = 1000$ draws of $\boldsymbol{y}$ from model (5.1) for each parameter setting. Tables 5.6, 5.7 and 5.8 show results for noticeable overdispersion, $\tau = 1$, for group-wise sample size of 10, 20 and 50, respectively. Tables 5.9 and 5.10 show results for low overdispersion, $\tau = 10$, for group-wise sample size of 10 and 20 respectively.

Table 5.6: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1) with noticeable overdispersion, $\tau = 1$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	10	0.1	−45.0	0.887	0.937	0.950	0.891	0.940	0.951
50	10	0.5	−25.0	0.877	0.941	0.936	0.889	0.936	0.953
50	10	0.8	−10.0	0.880	0.945	0.935	0.889	0.943	0.946
50	10	1.0	0.0	0.889	0.949	0.940	0.889	0.949	0.940
50	10	1.2	10.0	0.894	0.946	0.948	0.891	0.946	0.945
50	10	2.0	50.0	0.887	0.941	0.946	0.881	0.948	0.933
50	10	5.0	200.0	0.882	0.938	0.944	0.878	0.942	0.936
10	10	0.1	−9.0	0.840	0.925	0.915	0.816	0.890	0.926
10	10	0.5	−5.0	0.849	0.921	0.928	0.856	0.913	0.943
10	10	0.8	−2.0	0.849	0.916	0.933	0.855	0.917	0.938
10	10	1.0	0.0	0.857	0.929	0.928	0.857	0.929	0.928
10	10	1.2	2.0	0.868	0.933	0.935	0.865	0.936	0.929
10	10	2.0	10.0	0.875	0.934	0.941	0.857	0.942	0.915
10	10	5.0	40.0	0.869	0.935	0.934	0.882	0.959	0.923
10	10	10.0	90.0	0.878	0.944	0.934	0.886	0.951	0.935
1	10	0.1	−0.9	0.604	0.952	0.652	0.740	0.824	0.916
1	10	0.5	−0.5	0.758	0.880	0.878	0.760	0.850	0.910
1	10	0.8	−0.2	0.751	0.864	0.887	0.754	0.858	0.896
1	10	1.0	0.0	0.752	0.864	0.888	0.752	0.864	0.888
1	10	1.2	0.2	0.743	0.860	0.883	0.749	0.870	0.879
1	10	2.0	1.0	0.771	0.874	0.897	0.755	0.887	0.868
1	10	5.0	4.0	0.818	0.894	0.924	0.809	0.927	0.882
1	10	10.0	9.0	0.849	0.907	0.942	0.844	0.933	0.911

Table 5.7: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1), with noticeable overdispersion, $\tau = 1$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	20	0.1	−45.0	0.899	0.948	0.951	0.902	0.948	0.954
50	20	0.5	−25.0	0.893	0.944	0.949	0.898	0.946	0.952
50	20	0.8	−10.0	0.878	0.938	0.940	0.886	0.939	0.947
50	20	1.0	0.0	0.886	0.939	0.947	0.886	0.939	0.947
50	20	1.2	10.0	0.893	0.946	0.947	0.894	0.946	0.948
50	20	2.0	50.0	0.892	0.946	0.946	0.893	0.956	0.937
50	20	5.0	200.0	0.895	0.944	0.951	0.892	0.949	0.943
10	20	0.1	−9.0	0.885	0.956	0.929	0.882	0.939	0.943
10	20	0.5	−5.0	0.880	0.943	0.937	0.886	0.941	0.945
10	20	0.8	−2.0	0.875	0.940	0.935	0.871	0.939	0.932
10	20	1.0	0.0	0.875	0.942	0.933	0.875	0.942	0.933
10	20	1.2	2.0	0.893	0.948	0.945	0.889	0.948	0.941
10	20	2.0	10.0	0.891	0.953	0.938	0.898	0.958	0.940
10	20	5.0	40.0	0.882	0.945	0.937	0.881	0.945	0.936
10	20	10.0	90.0	0.879	0.947	0.932	0.875	0.939	0.936
1	20	0.1	−0.9	0.816	0.954	0.862	0.783	0.870	0.913
1	20	0.5	−0.5	0.813	0.915	0.898	0.795	0.889	0.906
1	20	0.8	−0.2	0.798	0.903	0.895	0.803	0.903	0.900
1	20	1.0	0.0	0.798	0.905	0.893	0.798	0.905	0.893
1	20	1.2	0.2	0.811	0.913	0.898	0.807	0.918	0.889
1	20	2.0	1.0	0.826	0.914	0.912	0.818	0.932	0.886
1	20	5.0	4.0	0.873	0.940	0.933	0.870	0.954	0.916
1	20	10.0	9.0	0.881	0.940	0.941	0.845	0.931	0.914

Table 5.8: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1), with noticeable overdispersion, $\tau = 1$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	50	0.1	−45.0	0.904	0.949	0.955	0.908	0.949	0.959
50	50	0.5	−25.0	0.906	0.949	0.957	0.905	0.947	0.958
50	50	0.8	−10.0	0.901	0.948	0.953	0.905	0.951	0.954
50	50	1.0	0.0	0.896	0.950	0.946	0.895	0.949	0.946
50	50	1.2	10.0	0.892	0.954	0.938	0.893	0.954	0.939
50	50	2.0	50.0	0.894	0.949	0.945	0.896	0.951	0.945
50	50	5.0	200.0	0.888	0.953	0.935	0.894	0.957	0.937
10	50	0.1	−9.0	0.886	0.955	0.931	0.882	0.944	0.938
10	50	0.5	−5.0	0.894	0.948	0.946	0.905	0.953	0.952
10	50	0.8	−2.0	0.893	0.951	0.942	0.891	0.951	0.940
10	50	1.0	0.0	0.901	0.958	0.943	0.901	0.958	0.943
10	50	1.2	2.0	0.907	0.958	0.949	0.912	0.959	0.953
10	50	2.0	10.0	0.903	0.953	0.950	0.905	0.957	0.948
10	50	5.0	40.0	0.903	0.955	0.948	0.906	0.955	0.951
10	50	10.0	90.0	0.909	0.956	0.953	0.906	0.955	0.951
1	50	0.1	−0.9	0.862	0.964	0.898	0.841	0.910	0.931
1	50	0.5	−0.5	0.857	0.937	0.920	0.850	0.925	0.925
1	50	0.8	−0.2	0.850	0.937	0.913	0.857	0.939	0.918
1	50	1.0	0.0	0.867	0.943	0.924	0.867	0.943	0.924
1	50	1.2	0.2	0.870	0.941	0.929	0.868	0.944	0.924
1	50	2.0	1.0	0.887	0.956	0.931	0.882	0.959	0.923
1	50	5.0	4.0	0.902	0.953	0.949	0.890	0.955	0.935
1	50	10.0	9.0	0.895	0.948	0.947	0.897	0.952	0.945

Table 5.9: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1), with low overdispersion, $\tau = 10$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	10	0.1	−45.0	0.866	0.945	0.921	0.807	0.901	0.906
50	10	0.5	−25.0	0.845	0.937	0.908	0.828	0.921	0.907
50	10	0.8	−10.0	0.857	0.938	0.919	0.847	0.933	0.914
50	10	1.0	0.0	0.868	0.944	0.924	0.867	0.944	0.923
50	10	1.2	10.0	0.869	0.942	0.927	0.860	0.944	0.916
50	10	2.0	50.0	0.858	0.938	0.920	0.861	0.944	0.917
50	10	5.0	200.0	0.865	0.938	0.927	0.859	0.943	0.916
10	10	0.1	−9.0	0.857	0.952	0.905	0.802	0.889	0.913
10	10	0.5	−5.0	0.828	0.922	0.906	0.805	0.901	0.904
10	10	0.8	−2.0	0.808	0.903	0.905	0.802	0.897	0.905
10	10	1.0	0.0	0.814	0.912	0.902	0.814	0.912	0.902
10	10	1.2	2.0	0.820	0.906	0.914	0.821	0.906	0.915
10	10	2.0	10.0	0.822	0.913	0.909	0.813	0.918	0.895
10	10	5.0	40.0	0.853	0.917	0.936	0.836	0.925	0.911
10	10	10.0	90.0	0.868	0.926	0.942	0.848	0.932	0.916
1	10	0.1	−0.9	0.604	0.971	0.633	0.874	0.908	0.966
1	10	0.5	−0.5	0.859	0.940	0.919	0.875	0.935	0.940
1	10	0.8	−0.2	0.857	0.940	0.917	0.861	0.934	0.927
1	10	1.0	0.0	0.864	0.934	0.930	0.864	0.934	0.930
1	10	1.2	0.2	0.871	0.935	0.936	0.869	0.936	0.933
1	10	2.0	1.0	0.862	0.925	0.937	0.867	0.948	0.919
1	10	5.0	4.0	0.866	0.918	0.948	0.827	0.925	0.902
1	10	10.0	9.0	0.875	0.917	0.958	0.815	0.919	0.896

Table 5.10: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1), with low overdispersion, $\tau = 10$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	20	0.1	−45.0	0.899	0.951	0.948	0.859	0.922	0.937
50	20	0.5	−25.0	0.879	0.937	0.942	0.884	0.932	0.952
50	20	0.8	−10.0	0.884	0.932	0.952	0.881	0.928	0.953
50	20	1.0	0.0	0.884	0.933	0.951	0.882	0.932	0.950
50	20	1.2	10.0	0.889	0.937	0.952	0.891	0.942	0.949
50	20	2.0	50.0	0.893	0.940	0.953	0.897	0.952	0.945
50	20	5.0	200.0	0.893	0.935	0.958	0.895	0.941	0.954
10	20	0.1	−9.0	0.888	0.952	0.936	0.822	0.888	0.934
10	20	0.5	−5.0	0.851	0.920	0.931	0.839	0.904	0.935
10	20	0.8	−2.0	0.833	0.906	0.927	0.833	0.900	0.933
10	20	1.0	0.0	0.843	0.913	0.930	0.843	0.913	0.930
10	20	1.2	2.0	0.849	0.916	0.933	0.847	0.917	0.930
10	20	2.0	10.0	0.864	0.918	0.946	0.858	0.917	0.941
10	20	5.0	40.0	0.870	0.921	0.949	0.874	0.936	0.938
10	20	10.0	90.0	0.883	0.928	0.955	0.888	0.939	0.949
1	20	0.1	−0.9	0.836	0.972	0.864	0.854	0.913	0.941
1	20	0.5	−0.5	0.871	0.944	0.927	0.870	0.932	0.938
1	20	0.8	−0.2	0.870	0.934	0.936	0.868	0.934	0.934
1	20	1.0	0.0	0.869	0.934	0.935	0.869	0.934	0.935
1	20	1.2	0.2	0.873	0.940	0.933	0.880	0.942	0.938
1	20	2.0	1.0	0.865	0.933	0.932	0.881	0.948	0.933
1	20	5.0	4.0	0.850	0.921	0.929	0.848	0.931	0.917
1	20	10.0	9.0	0.858	0.919	0.939	0.832	0.929	0.903

5.7 A gamma prior for the dispersion parameter

5.7.1 BUGS code and update parameters

The following model BUGS code represents the one-way model with negative binomial response of Equation (5.1), with $I = 2$ groups. In the following simulations the model of section 5.6.1 is used with the only difference that

```
r ~ unif(0,1000)
```

is replaced by

```
r ~ dgamma(0.001,0.001).
```

I.e., in contrast to the model in Section 5.6, a (conjugate), non-informative gamma prior is assumed for the parameter τ of the negative binomial distribution $\tau \sim gamma\,(0.001, 1000)$.

5.7.2 Detailed results

The technical details for running MCMC are the same as in Section 5.6.2. The estimated coverage probability is based on S=10000 random draws of $\boldsymbol{y}$ from the model in Equation (5.1) with $I = 2$, $\tau = 1$ and further parameters as indicated in the headers of the Tables. Tables 5.11 shows situations with high, intermediate and low mean abundance and group-wise sample size of 10.

Table 5.11: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1), with noticeable overdispersion, $\tau = 1$. Simulations are based on K=1000, S=10000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
50	10	0.1	−45.0	0.885	0.941	0.944	0.889	0.940	0.949
50	10	0.5	−25.0	0.888	0.945	0.943	0.891	0.938	0.952
50	10	0.8	−10.0	0.883	0.944	0.939	0.884	0.941	0.943
50	10	1.0	0.0	0.886	0.947	0.939	0.887	0.948	0.940
50	10	1.2	10.0	0.885	0.945	0.940	0.884	0.947	0.937
50	10	2.0	50.0	0.888	0.949	0.939	0.888	0.954	0.934
50	10	5.0	200.0	0.892	0.949	0.943	0.895	0.956	0.940
10	10	0.1	−9.0	0.881	0.952	0.930	0.886	0.939	0.947
10	10	0.5	−5.0	0.885	0.944	0.941	0.893	0.941	0.953
10	10	0.8	−2.0	0.886	0.943	0.943	0.889	0.941	0.948
10	10	1.0	0.0	0.886	0.944	0.942	0.886	0.944	0.942
10	10	1.2	2.0	0.888	0.943	0.945	0.886	0.944	0.942
10	10	2.0	10.0	0.888	0.944	0.944	0.887	0.952	0.936
10	10	5.0	40.0	0.883	0.944	0.939	0.883	0.950	0.933
10	10	10.0	90.0	0.889	0.948	0.941	0.887	0.950	0.937
1	10	0.1	0.9	0.598	0.973	0.625	0.887	0.939	0.948
1	10	0.5	0.5	0.862	0.941	0.920	0.867	0.921	0.946
1	10	0.8	0.2	0.862	0.931	0.931	0.863	0.923	0.940
1	10	1.0	0.0	0.865	0.928	0.938	0.865	0.927	0.938
1	10	1.2	−0.2	0.866	0.927	0.939	0.867	0.932	0.935
1	10	2.0	−1.0	0.872	0.928	0.944	0.871	0.942	0.928
1	10	5.0	−4.0	0.879	0.925	0.954	0.880	0.948	0.932
1	10	10.0	−9.0	0.886	0.929	0.957	0.887	0.946	0.941

5.8 Imposing weakly informative priors on the log-means

5.8.1 BUGS code and update parameters

The following model BUGS code represents the one-way model with negative binomial response of Equation (5.1), with $I = 2$ groups. The code shown in Section 5.6.1 is used with the only difference that

```
beta[p] ~ dnorm(0, 0.001)
```

is replaced by

```
beta[p] ~ dnorm(0, 0.1).
```

The technical details for running MCMC are the same as in Section 5.6.3. Note that in contrast to the model in Section 5.6, a weakly informative prior, `tau=0.1`, is imposed on the mean parameter on the log-scale, $\boldsymbol{\beta}$, i.e. $\beta_i \sim N(\mu = 0, \sigma^2 = 10)$.

5.8.2 Detailed results

The estimated coverage probability is based on S =1000 random draws of $\boldsymbol{y}$ from the model in Equation (5.1) with $I = 2$, $\tau = 1$ and further parameters as indicated in the headers of the tables. Table 5.12, and 5.13 show situations with intermediate and low mean abundance, for group-wise sample sizes of 10 and 20, respectively.

Table 5.12: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1) with noticeable overdispersion, $\tau = 1$, and a weakly informative prior imposed on the mean parameter on the log-scale, $\boldsymbol{\beta}$. Simulations based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
10	10	0	−9.000	0.845	0.918	0.927	0.815	0.881	1
10	10	0	−5.000	0.852	0.920	0.932	0.855	0.909	1
10	10	1	−2.000	0.842	0.913	0.929	0.843	0.909	1
10	10	1	0.000	0.855	0.927	0.928	0.855	0.927	1
10	10	1	2.000	0.865	0.933	0.932	0.863	0.934	1
10	10	2	10.000	0.868	0.930	0.938	0.857	0.946	1
10	10	5	40.000	0.866	0.938	0.928	0.867	0.963	1
10	10	10	90.000	0.873	0.946	0.927	0.873	0.961	1
1	10	0	−0.900	0.916	0.941	0.975	0.747	0.824	1
1	10	0	−0.500	0.777	0.878	0.899	0.773	0.852	1
1	10	1	−0.200	0.750	0.863	0.887	0.762	0.857	1
1	10	1	0.000	0.754	0.865	0.889	0.754	0.865	1
1	10	1	0.200	0.744	0.860	0.884	0.760	0.875	1
1	10	2	1.000	0.770	0.878	0.892	0.758	0.892	1
1	10	5	4.000	0.816	0.896	0.920	0.808	0.935	1
1	10	10	9.000	0.856	0.918	0.938	0.847	0.941	1

Table 5.13: Coverage probability of two-sided nominal 0.9 confidence intervals for the ratio and difference of means, assuming the negative binomial model in Equation (5.1) with noticeable overdispersion, $\tau = 1$ and a weakly informative prior imposed on the mean parameter on the log-scale, $\boldsymbol{\beta}$. Simulations are based on K=5000, S=1000.

				Ratio ρ			Difference δ		
μ_1	n_i	μ_2/μ_1	$\mu_2 - \mu_1$	CPts	CPl	CPu	CPts	CPl	CPu
10	10	0.1	−9.0	0.889	0.955	0.934	0.884	0.933	0.951
10	10	0.5	−5.0	0.879	0.943	0.936	0.888	0.939	0.949
10	10	0.8	−2.0	0.875	0.945	0.930	0.875	0.939	0.936
10	10	1.0	0.0	0.877	0.945	0.932	0.877	0.945	0.932
10	10	1.2	2.0	0.886	0.943	0.943	0.888	0.947	0.941
10	10	2.0	10.0	0.891	0.953	0.938	0.897	0.960	0.937
10	10	5.0	40.0	0.878	0.946	0.932	0.875	0.952	0.923
10	10	10.0	90.0	0.883	0.953	0.930	0.873	0.948	0.925
1	10	0.1	−0.9	0.873	0.938	0.935	0.780	0.861	0.919
1	10	0.5	−0.5	0.816	0.914	0.902	0.795	0.887	0.908
1	10	0.8	−0.2	0.797	0.902	0.895	0.801	0.900	0.901
1	10	1.0	0.0	0.804	0.906	0.898	0.804	0.906	0.898
1	10	1.2	0.2	0.814	0.915	0.899	0.807	0.917	0.890
1	10	2.0	1.0	0.834	0.920	0.914	0.820	0.934	0.886
1	10	5.0	4.0	0.876	0.943	0.933	0.867	0.957	0.910
1	10	10.0	9.0	0.879	0.944	0.935	0.847	0.935	0.912

Chapter 6

Simultaneous confidence intervals for ratios of means of negative binomials in the one-way layout

In this Chapter, simultaneous confidence intervals for ratios of means assuming a negative binomial response are investigated. Consider model (5.1) with $I > 2$, for example $I = 4$. Then, interest might be in comparisons to control defined as ratios as in Equations (4.9) and (4.13), Section 4.5. However, the methods described below are generally applicable when the parameter of interest can be defined as $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$.

6.1 Wald-type SCI for the ratio of means to the control mean

Fitting the model yields an estimate $\hat{\beta}$ for β and an estimate of the variance covariance matrix $\hat{\Sigma}$ of $\hat{\beta}$. In the case of mutually independent observations and a parameterization as outlined in (4.7) the off-diagonal elements of $\boldsymbol{\Sigma}$ are zero and only the I diagonal elements $\hat{\sigma}_i^2$ are of interest: $\hat{\boldsymbol{\sigma}}^2 = \text{diag}\left(\hat{\boldsymbol{\Sigma}}\right)$. Assuming asymptotic normality on the scale of β, one can construct simultaneous asymptotic $(1-\alpha)$

confidence intervals by using Equation 6.1.

$$\left[\hat{\theta}_m^l; \hat{\theta}_m^u\right] = \exp\left(\sum_{i=1}^{I} c_{mi}\hat{\beta}_i \pm z_{M,R,1-\alpha}^{two-sided}\sqrt{\sum_{i=1}^{I} c_{mi}^2\hat{\sigma}_i^2}\right) \tag{6.1}$$

The correlation matrix R with elements $\rho_{mm'}$ depends on known constants c_{mi} as well as on unknown parameters which have to be estimated.

$$\rho_{mm'} = \frac{\sum_{i=1}^{I} c_{im}c_{im'}\hat{\sigma}_i^2}{\sqrt{\sum_{i=1}^{I} c_{im}^2\hat{\sigma}_i^2 \sum_{i=1}^{I} c_{im'}^2\hat{\sigma}_i^2}} \tag{6.2}$$

As a special case, such confidence intervals are considered by Hothorn et al. (2008). However, their small sample properties for particular distributional assumptions have not been considered. Similar to the two-sample Wald type intervals considered in Section 5.3, this method will yield solutions practically equivalent to the parameter space, when the extreme event $y_n \equiv 0$ is observed for all n belonging to one group involved in the contrast. Less importantly, the extreme variance estimates resulting in that event disturb the estimation of the correlation structure in (6.2). As can be expected from the results shown in Section 5.3 and those in Gerhard and Schaarschmidt (2007), Wald type simultaneous confidence intervals show a conservative performance for small samples and abundances in unpublished simulation studies by (Gerhard, 2008, personal communication). The performance of such intervals is not considered here in detail.

6.2 SCI based on MCMC

One can translate the statistical model in Equation (5.1) for the case $I = 4$ into the `BUGS` model in Section 6.3.2 with non-informative priors imposed on all parameters. Note that the non-informative prior for the inverse dispersion parameter τ of the negative binomial distribution is gamma distribution, $\tau \sim gamma\,(0.001, 1000)$. Although the use of gamma priors is discouraged by a number of authors, the confidence intervals derived from the model in Section 6.3.2 performed uniformly better than those derived from models with a uniform prior $\tau \sim unif\,(0, 1000)$ (results not shown).

Following the methodology in Sections 4.4 and 4.5, simultaneous confidence intervals can be constructed. Following the Equations (4.8) and (4.9) yields ratios of mean abundances; following Equation (4.12) yields differences of mean abundances, for which interval construction is not straightforward based on the generalized linear model fits in Section 6.1 above.

6.3 MCMC derived SCI: Simulation study

As outlined in Section 1.4, simultaneous confidence intervals may be used for estimating the effect size, for testing hypotheses of non-inferiority vs. several standards, or for performing a proof of hazard. Main focus of the simulation study is on comparisons to control (Tables 6.1, 6.2, 6.3, 6.4). The simultaneous coverage probability of lower limits with nominal level 0.95 and of two-sided simultaneous confidence intervals with nominal level 0.95 was assessed. In order to expand the scope to multiple contrasts in general, also all pairwise comparisons as outlined in (4.19) and Williams type contrasts as defined in (4.20), Section 4.5, are considered for a limited number of settings (Tables 6.5, 6.6, 6.7, 6.8). In general, the ratios are defined by $\rho_m = \exp\left(\sum_{i=1}^{I} c_{mi}\beta_i\right)$ and the differences are defined by $\delta_m = \sum_{i=1}^{I} c_{im}\exp(\beta_i)$. As for the two-sample comparisons, high ($\cong 50$), intermediate ($\cong 10$) and low ($\cong 1$) mean abundances are considered, with $\exp(\beta_i) \in [0.1, 250]$. The inverse dispersion parameter τ of the negative binomial distribution was fixed at $\tau = 1$ in all the settings considered. Predominantly, balanced designs with group-wise sample size of 20 and 10 are considered. Moderately unbalanced designs are considered merely for the Williams contrast (Tables 6.7 and 6.8).

6.3.1 Summary of results

Tables 6.1 and 6.2 show simultaneous coverage probabilities of lower simultaneous 0.95 confidence limits for ratios and differences to control. For both parameters, the confidence intervals achieve close to nominal or slightly too low coverage prob-

abilities, when sample sizes are at least 20, or mean abundances are about 10 or greater. With mean abundances about 1 and group-wise sample size 10, the actual coverage probabilities of nominal 0.95 lower limits were as low as 0.91-0.92 for ratios and 0.89-0.93 for differences.

For two-sided 0.95 confidence limits (Tables 6.3 and 6.4), the results are similar, with an overall tendency of being too liberal. The coverage probability of the lower and upper bounds is close to, but usually below 0.975, with upper bounds being slightly more liberal in most settings. The most extreme deviations from the nominal level occurred for settings with small mean abundance and sample size 10: 0.75-0.93, 0.91-0.92 for ratio and difference, respectively. Most liberal results within settings with similar abundance were observed for extreme effect sizes ($\exp(\beta)' = (50, 50, 50, 250)$, and $\exp(\beta)' = (1, 1, 1, 0.1)$), due to markedly liberal upper bounds. Reasons for this might be that the number of updates was too low for this situation and the chosen initial values, such that the confidence intervals have been constructed based on joint distributions which did not have converged for the most extreme parameter. However, the technical parameters were chosen such that (for less extreme settings) simulated example data sets showed no signs of autocorrelation among consecutive values of the MCMC chains and Gewekes test did not reject the null-hypothesis of convergence. Moreover, applying Gewekes test on the simulation results did not lead to marked differences in the number of simulation runs which showed deviations from the null hypothesis (results not shown) among the settings considered in Tables 6.1 and 6.2.

For Tukey type contrasts (Tables 6.5 and 6.6) the simultaneous confidence intervals are even more liberal: For ratios, the coverage probabilities are 0.92-0.93 with sample size 10 and intermediate to high abundances, and 0.67-0.91 for the settings with low abundance. For differences, they range in 0.9-0.93 for intermediate to high abundances and 0.87-0.9 for low abundances. Again, most extreme deviations occur for the setting involving extremely low abundance: $\exp(\beta)' = (1, 1, 1, 0.1)$.

Williams contrasts (with parameters being pooled means over several β_i) show a more stable performance, comparable or better than that of the Dunnett con-

trasts (Tables 6.7 and 6.8 for ratios and differences). For all considered settings ($n_i \in [10, 35]$) and $\exp(\beta_i) \in [1, 250]$, observed coverage probabilities range between 0.94 and 0.95. Hence, also pooling contrasts in the presence of unbalanced settings may lead to simultaneous confidence intervals based on MCMC samples of the joint posterior, with acceptable frequentist properties.

6.3.2 BUGS code and update parameters

The one-way model with negative binomial response as defined in model (5.1), with $I = 4$ groups may lead to the following BUGS model:

```
model
{
for(n in 1:N)
{
X[n,1] <- X1[n]
X[n,2] <- X2[n]
X[n,3] <- X3[n]
X[n,4] <- X4[n]
Y[n] ~ dnegbin(pi[n], r)
pi[n] <- r/(r+mu[n])
mu[n] <- exp(eta[n])
eta[n] <- inprod(X[n,], beta[])
}
for(p in 1:P)
{
beta[p] ~ dnorm(0, 0.001)
muvec[p]<-exp(beta[p])
}
r ~ dgamma(0.001,0.001)
}
```

For running the model, the vectors `Y`, `X1`, `X2`, `X3`, `X4` and the integers `N` and `P` have to be provided along with initial values for the vector `beta` and the real `r`.

In the simulations, the model was updated 7000 times, with first 5000 updates discarded and every second value discarded in the remaining part of the chain, resulting in a sample of $K = 1000$ values from the joint distribution (Tables 6.3, 6.4, and 6.5). All estimated coverage probabilities are based on S=1000 simulation runs.

6.3.3 Detailed results

Lower 0.95 limits for Dunnett type contrasts

Table 6.1: Simultaneous coverage probability of lower 0.95 simultaneous confidence limits derived from MCMC with $I = 4$, group-wise sample size 10, markedly overdispersed data, $\tau = 1$, for Dunnett-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$	Difference $\boldsymbol{\delta}$
μ_1	μ_2	μ_3	μ_4	SCPl	SCPl
50.0	50.0	50.0	50.0	0.946	0.943
50.0	62.5	50.0	62.5	0.946	0.940
50.0	40.0	50.0	40.0	0.954	0.951
50.0	50.0	50.0	250.0	0.942	0.948
50.0	50.0	50.0	5.0	0.939	0.926
10.0	10.0	10.0	10.0	0.945	0.942
10.0	12.5	10.0	12.5	0.937	0.935
10.0	8.0	10.0	8.0	0.953	0.945
10.0	10.0	10.0	50.0	0.951	0.948
10.0	10.0	10.0	1.0	0.953	0.935
1.0	1.0	1.0	1.0	0.910	0.895
1.0	1.2	1.0	1.2	0.915	0.911
1.0	0.8	1.0	0.8	0.914	0.899
1.0	1.0	1.0	5.0	0.912	0.933
1.0	1.0	1.0	0.1	0.920	0.887

Table 6.2: Simultaneous coverage probability of lower 0.95 simultaneous confidence limits derived from MCMC with $I = 4$, group-wise sample size 20, markedly overdispersed data, $\tau = 1$, for Dunnett-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$	Difference $\boldsymbol{\delta}$
μ_1	μ_2	μ_3	μ_4	SCPl	SCPl
50.0	50.0	50.0	50.0	0.937	0.935
50.0	62.5	50.0	62.5	0.947	0.943
50.0	40.0	50.0	40.0	0.948	0.944
50.0	50.0	50.0	250.0	0.947	0.942
50.0	50.0	50.0	5.0	0.940	0.936
10.0	10.0	10.0	10.0	0.945	0.940
10.0	12.5	10.0	12.5	0.955	0.953
10.0	8.0	10.0	8.0	0.954	0.943
10.0	10.0	10.0	50.0	0.944	0.947
10.0	10.0	10.0	1.0	0.955	0.942
1.0	1.0	1.0	1.0	0.944	0.935
1.0	1.2	1.0	1.2	0.942	0.938
1.0	0.8	1.0	0.8	0.940	0.936
1.0	1.0	1.0	5.0	0.950	0.949
1.0	1.0	1.0	0.1	0.950	0.943

Two-sided 0.95 intervals for Dunnett type contrasts

Table 6.3: Simultaneous coverage probability (1000 simulations) of two-sided nominal 0.95 SCI derived from MCMC with $I = 4$, $n_i = 10$, $\tau = 1$, for Dunnett-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50.0	50.0	50.0	50.0	0.938	0.971	0.967	0.928	0.970	0.958
50.0	62.5	50.0	62.5	0.944	0.976	0.968	0.941	0.974	0.967
50.0	40.0	50.0	40.0	0.949	0.977	0.972	0.944	0.976	0.968
50.0	50.0	50.0	250.0	0.925	0.969	0.956	0.920	0.971	0.949
50.0	50.0	50.0	5.0	0.944	0.974	0.970	0.942	0.975	0.967
10.0	10.0	10.0	10.0	0.941	0.971	0.970	0.938	0.970	0.968
10.0	12.5	10.0	12.5	0.945	0.975	0.970	0.942	0.975	0.967
10.0	8.0	10.0	8.0	0.951	0.975	0.976	0.943	0.972	0.971
10.0	10.0	10.0	50.0	0.936	0.968	0.968	0.936	0.972	0.964
10.0	10.0	10.0	1.0	0.937	0.976	0.961	0.938	0.978	0.960
1.0	1.0	1.0	1.0	0.919	0.961	0.958	0.915	0.965	0.950
1.0	1.2	1.0	1.2	0.927	0.965	0.962	0.923	0.970	0.953
1.0	0.8	1.0	0.8	0.914	0.960	0.954	0.909	0.960	0.949
1.0	1.0	1.0	5.0	0.927	0.969	0.958	0.922	0.975	0.946
1.0	1.0	1.0	0.1	0.754	0.974	0.780	0.923	0.970	0.953

Table 6.4: Simultaneous coverage probability (1000 simulations) of two-sided 0.95 SCI derived from MCMC with $I = 4$, $n_i = 20$, $\tau = 1$, for Dunnett-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50.0	50.0	50.0	50.0	0.946	0.971	0.975	0.944	0.969	0.975
50.0	62.5	50.0	62.5	0.949	0.973	0.976	0.942	0.975	0.967
50.0	40.0	50.0	40.0	0.946	0.975	0.971	0.945	0.975	0.970
50.0	50.0	50.0	250.0	0.947	0.967	0.980	0.947	0.970	0.977
50.0	50.0	50.0	5.0	0.947	0.970	0.977	0.948	0.974	0.974
10.0	10.0	10.0	10.0	0.944	0.977	0.967	0.940	0.979	0.961
10.0	12.5	10.0	12.5	0.948	0.977	0.971	0.944	0.978	0.966
10.0	8.0	10.0	8.0	0.950	0.980	0.970	0.945	0.978	0.967
10.0	10.0	10.0	50.0	0.949	0.976	0.973	0.946	0.979	0.967
10.0	10.0	10.0	1.0	0.950	0.980	0.970	0.946	0.977	0.969
1.0	1.0	1.0	1.0	0.938	0.970	0.968	0.935	0.971	0.964
1.0	1.2	1.0	1.2	0.940	0.974	0.966	0.934	0.975	0.959
1.0	0.8	1.0	0.8	0.931	0.967	0.964	0.926	0.969	0.957
1.0	1.0	1.0	5.0	0.950	0.978	0.972	0.947	0.979	0.968
1.0	1.0	1.0	0.1	0.827	0.973	0.853	0.928	0.971	0.957

Two-sided 0.95 intervals for Tukey type contrasts

Table 6.5: Simultaneous coverage probability (1000 simulations) of two-sided nominal 0.95 SCI derived from MCMC with $I = 4$, $n_i = 10$, $\tau = 1$, for Tukey-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50.0	50.0	50.0	50.0	0.936	0.964	0.966	0.913	0.950	0.955
50.0	62.5	50.0	62.5	0.942	0.969	0.968	0.922	0.957	0.959
50.0	40.0	50.0	40.0	0.929	0.960	0.960	0.900	0.940	0.946
50.0	50.0	50.0	250.0	0.941	0.966	0.970	0.926	0.960	0.966
50.0	50.0	50.0	5.0	0.926	0.959	0.962	0.935	0.959	0.965
10.0	10.0	10.0	10.0	0.923	0.963	0.950	0.902	0.949	0.942
10.0	12.5	10.0	12.5	0.922	0.962	0.951	0.910	0.957	0.942
10.0	8.0	10.0	8.0	0.932	0.966	0.959	0.917	0.956	0.949
10.0	10.0	10.0	50.0	0.936	0.964	0.963	0.928	0.963	0.960
10.0	10.0	10.0	1.0	0.927	0.965	0.957	0.919	0.950	0.952
1.0	1.0	1.0	1.0	0.899	0.947	0.941	0.876	0.930	0.932
1.0	1.2	1.0	1.2	0.903	0.948	0.941	0.884	0.940	0.930
1.0	0.8	1.0	0.8	0.895	0.945	0.935	0.874	0.932	0.924
1.0	1.0	1.0	5.0	0.914	0.948	0.951	0.900	0.955	0.939
1.0	1.0	1.0	0.1	0.673	0.959	0.702	0.894	0.926	0.946

Table 6.6: Simultaneous coverage probability (1000 simulations) of two-sided nominal 0.95 SCI derived from MCMC with $I = 4$, $n_i = 20$, $\tau = 1$, for Tukey-type contrasts for the ratio and difference of means of negative binomial samples.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50.0	50.0	50.0	50.0	0.938	0.969	0.966	0.931	0.963	0.962
50.0	62.5	50.0	62.5	0.935	0.970	0.959	0.926	0.966	0.954
50.0	40.0	50.0	40.0	0.929	0.962	0.963	0.920	0.956	0.955
50.0	50.0	50.0	250.0	0.935	0.963	0.967	0.940	0.967	0.969
50.0	50.0	50.0	5.0	0.938	0.966	0.967	0.933	0.960	0.962
10.0	10.0	10.0	10.0	0.942	0.976	0.960	0.931	0.969	0.954
10.0	12.5	10.0	12.5	0.947	0.973	0.969	0.940	0.969	0.966
10.0	8.0	10.0	8.0	0.949	0.976	0.967	0.938	0.965	0.962
10.0	10.0	10.0	50.0	0.946	0.970	0.968	0.941	0.971	0.966
10.0	10.0	10.0	1.0	0.931	0.969	0.961	0.934	0.966	0.957
1.0	1.0	1.0	1.0	0.923	0.957	0.956	0.910	0.952	0.946
1.0	1.2	1.0	1.2	0.940	0.968	0.960	0.930	0.963	0.954
1.0	0.8	1.0	0.8	0.934	0.971	0.955	0.928	0.964	0.953
1.0	1.0	1.0	5.0	0.933	0.959	0.967	0.938	0.970	0.964
1.0	1.0	1.0	0.1	0.827	0.972	0.849	0.918	0.957	0.948

Two-sided 0.95 intervals for Williams type contrasts

Table 6.7: Simultaneous coverage probability of SCI derived from MCMC with $I = 4$, with moderately unbalanced group-wise sample sizes $n_1, ..., n_4$, and $\tau = 1$, for Williams-type contrasts defined at the ratio $\boldsymbol{\rho}$ of means $\mu_i = \exp(\beta_i)$ of negative binomial samples.

								Ratio $\boldsymbol{\rho}$		
μ_1	μ_2	μ_3	μ_4	n_1	n_2	n_3	n_4	SCPts	SCPl	SCPu
50	50.0	50	50.0	35	15	15	15	0.951	0.978	0.973
50	62.5	50	62.5	20	10	20	30	0.942	0.966	0.976
50	40.0	50	40.0	20	10	20	30	0.936	0.964	0.972
50	50.0	50	250.0	30	20	20	10	0.944	0.971	0.973
50	50.0	50	5.0	10	20	20	30	0.944	0.974	0.970
10	10.0	10	10.0	35	15	15	15	0.941	0.967	0.974
10	12.5	10	12.5	20	10	20	30	0.951	0.973	0.978
10	8.0	10	8.0	20	10	20	30	0.943	0.970	0.973
10	10.0	10	50.0	30	20	20	10	0.951	0.974	0.977
10	10.0	10	1.0	10	20	20	30	0.933	0.971	0.962

Table 6.8: Simultaneous coverage probability of SCI derived from MCMC with $I = 4$, with moderately unbalanced group-wise sample sizes $n_1, ..., n_4$, and $\tau = 1$, for Williams-type contrasts defined at the difference $\boldsymbol{\delta}$ of means $\mu_i = \exp(\beta_i)$ of negative binomial samples.

								Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	n_1	n_2	n_3	n_4	SCPts	SCPl	SCPu
50	50.0	50	50.0	35	15	15	15	0.949	0.980	0.969
50	62.5	50	62.5	20	10	20	30	0.945	0.971	0.974
50	40.0	50	40.0	20	10	20	30	0.932	0.962	0.970
50	50.0	50	250.0	30	20	20	10	0.950	0.975	0.975
50	50.0	50	5.0	10	20	20	30	0.945	0.976	0.969
10	10.0	10	10.0	35	15	15	15	0.941	0.968	0.973
10	12.5	10	12.5	20	10	20	30	0.951	0.975	0.976
10	8.0	10	8.0	20	10	20	30	0.943	0.973	0.970
10	10.0	10	50.0	30	20	20	10	0.951	0.974	0.977
10	10.0	10	1.0	10	20	20	30	0.947	0.978	0.969

Chapter 7

SCI for ratios of means in a number of simple mixed models assuming overdispersed count data

7.1 Estimation in hierarchical models using MCMC

In Bayesian modeling, there is no clear distinction of fixed and random effect models, since all parameters influencing the observable data are again modeled as random quantities derived from hyper distributions at a lower level (e.g. Clayton, 1996). In this way, also overdispersion might be modeled by a random effect on the scale of individual observations.

In a model with three levels of effects (Gelman et al., 2004):

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{\Sigma}_y &\sim N\left(\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\Sigma}_y\right), \\ \boldsymbol{\beta}|\boldsymbol{X}_\beta,\boldsymbol{\xi},\boldsymbol{\Sigma}_\beta &\sim N\left(\boldsymbol{X}_\beta\boldsymbol{\xi},\boldsymbol{\Sigma}_\beta\right), \\ \boldsymbol{\xi}|\boldsymbol{\xi}_0,\boldsymbol{\Sigma}_\xi &\sim N\left(\boldsymbol{\xi}_0,\boldsymbol{\Sigma}_\xi\right), \end{aligned} \tag{7.1}$$

the first part models the likelihood of the observable data $\boldsymbol{y}$ given the unknown, unobservable parameters $\boldsymbol{\beta}$ for mean and variance $\boldsymbol{\Sigma}_y$ and the assumption of a

Gaussian distribution, with the design matrix $\boldsymbol{X}$ describing the structure of the data with respect to $\boldsymbol{\beta}$. The second part models the prior knowledge or assumption on the distribution of the unobservable parameter of means, $\boldsymbol{\beta}$. Here, $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}_\beta$ model the mean, variance and covariance of $\boldsymbol{\beta}$, given an imposed structure $\boldsymbol{X}_\beta$ for the elements of $\boldsymbol{\beta}$. The third level models the parameters of the second level as a hyper distribution. Multiple levels of nesting or complex assumptions on the structure of effects add levels to the hierarchy or add complexity to design matrices $\boldsymbol{X}$ or the variance parameters $\boldsymbol{\Sigma}$.
Practically, adding levels to hierarchical models increases the computational burden in the update process of the Gibbs Sampler, increases the autocorrelation of the samples drawn from the posterior and thence slows down convergence. Technical strategies to increase the convergence ('hierarchical centering') are discussed e.g. by Gelfand et al. (1995).

The hierarchical models considered in detail in the literature often exhibit only simple treatment structures, e.g. linear regression problems. Thence only simple inferential problems follow for the 'fixed effects', i.e., the mean parameters on the first level (Zhao et al., 2006; Spiegelhalter et al., 2007). The problems of factorial treatment structures are rarely considered explicitly, e.g. by Clayton (1996) and Nobile and Green (2000).

7.2 Formal definition of the models considered

In the following, a number of models is formally introduced, which will later on be used in simulation studies. Throughout this chapter it is assumed that the fixed effects design matrix X is formulated without intrinsic aliasing of treatment levels (Clayton, 1996), i.e., with dummy coding of treatment levels chosen in a way that the model is not overspecified. Further, in the Equations (7.3) and (7.4), the non-informative priors are omitted. Hence, for all parameters for which no distributional assumption is stated in the following equations, an appropriate non-informative prior can be found in the `BUGS` code preceding the detailed results.

In order to be close to usual definitions of generalized linear mixed models in the frequentists sense (McCulloch and Searle, 2001), the following general notation is used:

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \tag{7.2}$$

Here, $\boldsymbol{\eta}$ is an $(N \times 1)$ vector of expectations on the link scale, with $n = 1, ..., N$ being the index of the observations. $\boldsymbol{X}$ is an $(N \times I)$ design matrix for those effects for which inferential interest is in each particular element, i.e., the design matrix of the fixed effects. The $(1 \times I)$ vector $\boldsymbol{\beta}$ with elements β_i, $i = 1, ..., I$ then is defined by the choice of $\boldsymbol{X}$. Since major interest is in general multiple comparisons among treatment groups, $\boldsymbol{X}$ will be defined as a outlined in Section 4.5, and illustrated in Equation (4.7).

Independent of the choice of $\boldsymbol{X}$, the choice of the $(N \times J)$ matrix $\boldsymbol{Z}$ defines the meaning of the $(1 \times J)$ vector $\boldsymbol{\gamma}$, in which there is no primary inferential interest, and which might be called the random effects part of the model. In all models considered, $\boldsymbol{Z}$ contains the information in which clusters the individual observations in $\boldsymbol{Y}$ were arranged. For all the simple hierarchical models introduced in the following sections, it is assumed that there is only one level of hierarchy in clustering, either in a simple multi-year or multi-location trial or in a simple repeated measurement design. Thence, the $\boldsymbol{Z}$ may again have the structure of a cell means model, allowing all elements of $\boldsymbol{\gamma}$, γ_j, to be modeled identically $\gamma_j \sim N(0, \sigma_u^2)$. Such simple random effects might be modeled in a computationally more efficient way. However, the chosen notation allows generalization to more complicated models.

Finally, the $(N \times 1)$ vector $\boldsymbol{\epsilon}$ with elements ϵ_n, $n = 1, ..., N$ models variability in the response $\boldsymbol{Y}$ at the level of individual observations which is not taken into account by variance of the distribution assumed for $\boldsymbol{Y}$. In the considered models, $\boldsymbol{\epsilon}$ models overdispersion of $\boldsymbol{Y}$ if $Y_n \sim Pois(\mu_n)$ is assumed.

In the following sections, three particular models are defined, which are used in simulation studies in Sections 7.3.1, 7.4.1, and 7.5.1 to assess the frequentist simultaneous

coverage probability of simultaneous credible intervals for linear combinations of $\boldsymbol{\beta}$. All models assume that the data can sufficiently described by a linear model on the link scale, using the log-link, i.e. $\log(\boldsymbol{\mu}) = \boldsymbol{\eta}$.

7.3 A simple hierarchical model with overdispersed Poisson response

Assume a trial where data are obtained from J clusters (e.g. years or locations), indexed by $j = 1, ..., J$. Within each cluster, I treatments are randomly assigned to the observational units, with treatments indexed by $i = 1, ..., I$. The total number of observations is denoted N, with the observations indexed by $n = 1, ..., N$. The treatment structure is represented in a $(N \times I)$ matrix $\boldsymbol{X}$. The affiliation of observations to the clusters are represented by a $(N \times J)$ matrix $\boldsymbol{Z}$. Note, that in this simple model, the effects of clusters and treatments are assumed to be additive on the log-scale. Denoting the elements of $\boldsymbol{X}$ as x_{ni} and the elements of $\boldsymbol{Z}$ as z_{nj}, the parameter vector $\boldsymbol{\beta}$ with elements β_i, $i = 1, ..., I$ models the means of treatments and the parameter $\boldsymbol{\gamma}$ with elements γ_i, $i = 1, ..., J$ models the variability between the clusters. Primary interest is in $\boldsymbol{\beta}$, for which the absence of prior information is assumed. There is no interest in the particular elements of $\boldsymbol{\gamma}$. Rather, the aim is to model the variability that is introduced into Y by the different clusters.

$$\begin{aligned} Y_n &\sim Pois(\mu_n) \\ \mu_n &= \exp(\eta_n) \\ \eta_n &= \beta_0 + \textstyle\sum_{i=1}^{I} x_{ni}\beta_i + \sum_{j=1}^{J} z_{nj}\gamma_j + \epsilon_n \\ \gamma_j &\sim N(0, \sigma_u^2) \\ \epsilon_n &\sim N(0, \sigma_e^2) \end{aligned} \tag{7.3}$$

The term ϵ_n models variability on the scale of individual measurements. Note that $\boldsymbol{Z\gamma}$ here just models the effects of clusters to be additive to the treatment effects modeled by $\boldsymbol{X\beta}$. This means that the model in (PoisMMod) assumes the absence of a cluster-treatment interaction.

7.3.1 Simulation study

In a simulation study, the simultaneous coverage probabilities of two-sided nominal 0.95 confidence intervals are assessed for a design with $I = 4$ treatments, $J = 10$ clusters and three replications within each cluster. Nine settings with $\mu_n \in [0.5, 250]$ are considered, once with extra-Poisson variability mainly emerging from dissimilarity among clusters $\sigma_u = 0.5$, $\sigma_e = 0.1$ (Table 7.1), and once with extra-Poisson variability mainly present at the level of individual observations $\sigma_u = 0.5$, $\sigma_e = 0.1$ (Table 7.2).

7.3.2 Summary of results

For eight of nine settings (Tables 7.1 and 7.2), the coverage probability is very close to the nominal level or slightly below. However, for the setting $\exp(\boldsymbol{\beta}) = (50, 50, 50, 250)$, the confidence intervals for the difference to control violate the nominal level severely for the upper limit. Apparently, the posterior did not converge for the chosen initial values $\exp(\boldsymbol{\beta}) = (1, 1, 1, 1)$ and the 15000 updates with first 5000 discarded. Such a situation is less likely to occur when sample estimates are used as initial values.

7.3.3 BUGS code and update parameters

A representation of the model (7.3) in the BUGS language for $I = 4$ and $J = 10$ is:

```
model{
for(i.obs in 1:n.obs)
{
X[1,i.obs] <- X1[i.obs]
X[2,i.obs] <- X2[i.obs]
X[3,i.obs] <- X3[i.obs]
X[4,i.obs] <- X4[i.obs]
Z[1,i.obs] <- Z1[i.obs]
```

```
Z[2,i.obs] <- Z2[i.obs]
Z[3,i.obs] <- Z3[i.obs]
Z[4,i.obs] <- Z4[i.obs]
Z[5,i.obs] <- Z5[i.obs]
Z[6,i.obs] <- Z6[i.obs]
Z[7,i.obs] <- Z7[i.obs]
Z[8,i.obs] <- Z8[i.obs]
Z[9,i.obs] <- Z9[i.obs]
Z[10,i.obs] <- Z10[i.obs]
epsi[i.obs] ~ dnorm(0, tau.e)
eta[i.obs] <- inprod(X[,i.obs], beta[])
 + inprod(Z[,i.obs],gamma[]) + epsi[i.obs]
log(mu[i.obs]) <- eta[i.obs]
Y[i.obs] ~ dpois(mu[i.obs])
}
for(i.treat in 1:n.treat)
{
beta[i.treat] ~ dnorm(0, 0.001)
}
for(i.year in 1:n.year)
{
gamma[i.year] ~ dnorm(0, tau.year)
}
tau.e ~ dgamma(0.001,0.001)
tau.year ~ dgamma(0.001, 0.001)
sigma.year <- 1/tau.year
}
```

The vectors `X1,...,X4`, `Z1,...,Z10`, and the numerics `n.obs`, `n.treat`, `n.year` have to be provided as input data. Initial values need to be provided for the vectors `beta[]`, `gamma[]`, `epsi[]`, and the numerics `tau.year`, `tau.e`.

7.3.4 Detailed results

The simulations are based on MCMC chains with 15000 updates, first 5000 updates discarded, and 9 out of 10 updates discarded in the remaining part of the chain, resulting in a sample of K=1000 values from the joint posterior. Estimated coverage probabilities are based on S=1000 simulation runs.

Table 7.1: Simultaneous coverage probabilities of nominal 0.95 two-sided confidence intervals for ratios $\boldsymbol{\rho}$ and differences $\boldsymbol{\delta}$ to control, following a hierarchical model assuming $Pois(\mu)$ response, $I = 4$ treatments, $J = 10$ clusters, and 3 replications of each treatments within each cluster. The variance parameter on the cluster level was chosen $\sigma_u = 0.5$, the variance parameter on the observation level (overdispersion parameter) was chosen $\sigma_e = 0.1$

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50	50	50	50.0	0.942	0.966	0.976	0.941	0.966	0.975
50	50	50	250.0	0.955	0.972	0.983	0.310	0.979	0.318
50	50	50	5.0	0.935	0.972	0.963	0.924	0.949	0.975
10	10	10	10.0	0.957	0.978	0.979	0.957	0.980	0.977
10	10	10	50.0	0.938	0.965	0.973	0.942	0.971	0.971
10	10	10	1.0	0.949	0.980	0.969	0.951	0.979	0.972
5	5	5	5.0	0.944	0.970	0.974	0.942	0.969	0.973
5	5	5	25.0	0.950	0.970	0.980	0.946	0.968	0.977
5	5	5	0.5	0.953	0.984	0.969	0.941	0.975	0.966

Table 7.2: Simultaneous coverage probabilities of nominal 0.95 two-sided confidence intervals, following a hierarchical model assuming $Pois(\mu)$ response, with $I = 4$ treatments, $J = 10$ clusters, and 3 replications of each treatments within each cluster. The variance parameter on the cluster level was chosen $\sigma_u = 0.1$, the variance parameter on the observation level (overdispersion) was chosen $\sigma_e = 0.5$

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50	50	50	50.0	0.953	0.974	0.979	0.951	0.975	0.976
50	50	50	250.0	0.948	0.974	0.974	0.260	0.982	0.265
50	50	50	5.0	0.959	0.985	0.974	0.937	0.962	0.975
10	10	10	10.0	0.945	0.973	0.972	0.944	0.975	0.969
10	10	10	50.0	0.943	0.975	0.968	0.945	0.979	0.966
10	10	10	1.0	0.924	0.963	0.961	0.944	0.971	0.973
5	5	5	5.0	0.939	0.971	0.968	0.936	0.971	0.965
5	5	5	25.0	0.941	0.964	0.977	0.936	0.966	0.970
5	5	5	0.5	0.936	0.962	0.974	0.938	0.964	0.974

7.4 A simple hierarchical model assuming negative binomial response

Another common approach to model overdispersion of count data is the negative binomial distribution:

$$\begin{aligned} Y_n &\sim NB\left(\mu_n, \tau\right) \\ \mu_n &= \exp\left(\eta_n\right) \\ \eta_n &= \sum_{i=1}^{I} x_{ni}\beta_i + \sum_{j=1}^{J} z_{nj}\gamma_j \\ \gamma_j &\sim N\left(0, \sigma_u^2\right) \end{aligned} \tag{7.4}$$

Here, the parameter τ models the overdispersion in dependence of μ_n as a common function for all treatments and clusters.

7.4.1 Simulation study

In the simulation study, the coverage probability for simultaneous confidence intervals was assessed for comparisons to control among $I = 4$ treatments in a design with $J = 10$ clusters and three replications of each treatment within each cluster. Tables 7.3, 7.4, and 7.5 show the simultaneous coverage probabilities for lower 0.95 simultaneous confidence limits, two-sided nominal 0.9, and twos-sided nominal 0.95 SCI, respectively. The fixed population means $\exp\left(\beta_i\right)$ range in $[0.5, 250]$, but are mostly fixed at 5, 10 and 50. The parameter modeling the variability were fixed at $\sigma_u = 0.5$ and $\tau = 1$.

7.4.2 Summary of results

For both the ratio and the difference, the observed simultaneous coverage probabilities are close to the nominal level, and only for one among the 27 considered settings the observed coverage probability is significantly lower than the nominal level. The probabilities to exclude the true parameter is about equally distributed to the upper and lower limits.

7.4.3 BUGS code and update parameters

A representation of the model (7.4) in the BUGS language for $I = 4$ and $J = 10$ is:

```
model{
for(i.obs in 1:n.obs)
{
X[1,i.obs] <- X1[i.obs]
X[2,i.obs] <- X2[i.obs]
X[3,i.obs] <- X3[i.obs]
X[4,i.obs] <- X4[i.obs]
Z[1,i.obs] <- Z1[i.obs]
Z[2,i.obs] <- Z2[i.obs]
Z[3,i.obs] <- Z3[i.obs]
Z[4,i.obs] <- Z4[i.obs]
Z[5,i.obs] <- Z5[i.obs]
Z[6,i.obs] <- Z6[i.obs]
Z[7,i.obs] <- Z7[i.obs]
Z[8,i.obs] <- Z8[i.obs]
Z[9,i.obs] <- Z9[i.obs]
Z[10,i.obs] <- Z10[i.obs]
eta[i.obs] <- inprod(X[,i.obs], beta[]) + inprod(Z[,i.obs],gamma[])
mu[i.obs] <- exp(eta[i.obs])
pi[i.obs] <- r/(r + mu[i.obs])
Y[i.obs] ~ dnegbin(pi[i.obs], r)
}
for(i.treat in 1:n.treat)
{
beta[i.treat] ~ dnorm(0, 0.001)
}
for(i.year in 1:n.year)
{
```

```
gamma[i.year] ~ dnorm(0, tau.year)
}
r ~ dgamma(0.001,0.001)
tau.year ~ dgamma(0.001, 0.001)
}
```

The vectors `X1,...,X4`, `Z1,...,Z10`, and the numerics `n.obs`, `n.treat`, `n.year` have to be provided as input data. Initial values need to provided for the terms `beta[]`, `gamma[]`, `r`, `tau.year`.

7.4.4 Detailed results

In each of 1000 simulations per parameter setting, a chain of 12000 updates was run, with first 2000 values in the chains discarded, and 1 out of 10 values retained after thinning, resulting in a sample of $K = 1000$ values from the joint posterior. The estimated coverage probabilities are based on S=1000 simulation runs.

Table 7.3: Simultaneous coverage probability of nominal lower 0.95 confidence limits following a hierarchical model assuming a $NB(\mu_i, \tau)$ response with common negative binomial dispersion parameter $\tau = 1$, with $I = 4$ treatments, $J = 10$ clusters, and 3 replications of each treatments within each cluster. The variance parameter on the cluster level was $\sigma_u = 0.5$.

				Ratio $\boldsymbol{\rho}$	Difference $\boldsymbol{\delta}$
μ_1	μ_2	μ_3	μ_4	SCPl	SCPl
50	50	50	50.0	0.951	0.948
50	50	50	250.0	0.946	0.945
50	50	50	5.0	0.951	0.947
10	10	10	10.0	0.954	0.950
10	10	10	50.0	0.940	0.939
10	10	10	1.0	0.949	0.941
5	5	5	5.0	0.952	0.948
5	5	5	25.0	0.952	0.953
5	5	5	0.5	0.961	0.955

Table 7.4: Simultaneous coverage probability for nominal 0.9 two-sided confidence intervals following a hierarchical model assuming a $NB(\mu_i, \tau)$ response with common dispersion parameter $\tau = 1$, with $I = 4$ treatments, $J = 10$ clusters, and 3 replications of each treatment within each cluster. The variance parameter on the cluster level was σ_u=0.5.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50	50	50	50.0	0.889	0.949	0.940	0.885	0.949	0.936
50	50	50	250.0	0.904	0.960	0.944	0.897	0.961	0.935
50	50	50	5.0	0.899	0.953	0.946	0.899	0.957	0.942
10	10	10	10.0	0.896	0.949	0.947	0.890	0.949	0.941
10	10	10	50.0	0.874	0.943	0.931	0.872	0.944	0.927
10	10	10	1.0	0.897	0.952	0.944	0.880	0.949	0.931
5	5	5	5.0	0.891	0.948	0.943	0.886	0.948	0.938
5	5	5	25.0	0.898	0.954	0.944	0.900	0.957	0.943
5	5	5	0.5	0.888	0.956	0.932	0.894	0.960	0.934

Table 7.5: Simultaneous coverage probability for nominal 0.95 two-sided confidence intervals following a hierarchical model assuming a $NB(\mu_i, \tau)$ response with common dispersion parameter $\tau = 1$, with I=4 treatments, J=10 clusters, and 3 replications of each treatment within each cluster. The variance parameter on the cluster level was $\sigma_u = 0.5$.

				Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
μ_1	μ_2	μ_3	μ_4	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50	50	50	50.0	0.952	0.982	0.970	0.951	0.983	0.968
50	50	50	250.0	0.954	0.982	0.972	0.944	0.979	0.965
50	50	50	5.0	0.944	0.978	0.966	0.947	0.976	0.971
10	10	10	10.0	0.947	0.970	0.977	0.943	0.970	0.973
10	10	10	50.0	0.943	0.969	0.974	0.939	0.970	0.969
10	10	10	1.0	0.941	0.975	0.966	0.945	0.976	0.969
5	5	5	5.0	0.947	0.975	0.972	0.940	0.972	0.968
5	5	5	25.0	0.954	0.980	0.974	0.948	0.977	0.971
5	5	5	0.5	0.938	0.978	0.960	0.946	0.979	0.967

7.5 A simple model for repeated measurements with overdispersed Poisson response

Let N denote the total number of observations with index $n = 1, ..., N$, let J denote the number of treatments $j = 1, ..., J$ and let H denote the number of randomized experimental units or clusters, $h = 1, ..., H$. Further, assume that within each experimental unit h, a number of T repeated measures are taken. The population effects of treatments and time, as well as their interaction is of primary interest in statistical inference. Then, $\boldsymbol{Y}$ is an $(N \times 1)$ vector of all observations (with $N = HT$), $\boldsymbol{X}$ is an $(N \times I)$, matrix with $I = JT$, containing the crossed treatment and time effects as dummy-coded variables and $\boldsymbol{Z}$ is a $(N \times H)$ matrix containing the information which observations of $\boldsymbol{Y}$ belong to the same experimental unit by H dummy coded variables. Let y_n, x_{ni}, and z_{nh} denote the elements of $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$, respectively.

$$\begin{aligned} Y_n &\sim Pois\left(\mu_n\right) \\ \mu_n &= \exp\left(\eta_n\right) \\ \eta_n &= \textstyle\sum_{i=1}^{I} x_{ni}\beta_i + \sum_{h=1}^{H} z_{nh}\gamma_h + \epsilon_n \\ \gamma_h &\sim N\left(0, \sigma_h^2\right) \\ \epsilon_n &\sim N\left(0, \sigma_e^2\right) \end{aligned} \tag{7.5}$$

In this model, the parameters β_i, $i = 1, ..., I$, $I = JT$ model the treatment-time-interaction on the log-scale, the parameters γ_h model the mean differences between the experimental units on the log-scale and the quantities ϵ_n model the extra-Poisson variability of the response. The correlation among measurements within each subject is assumed equal, i.e., a compound symmetry model (assuming that the individual measurements are exchangeable) is applied.

Note, that with this parametrization the computational burden becomes relatively high since sums over a large number of (unnecessary) products have to be calculated which contain mainly zero elements. In difference to the model in Equations (7.3) and (7.4), here it is assumed that the random effect is clustered within the subjects

and hence within the treatments which are unique for each subject whereas in (7.3) and (7.4) the treatments are crossed with the random effects, i.e., all treatments may occur within each cluster.

7.5.1 Simulation study

The Tables 7.8 and 7.9 in Section 7.5.1) show simultaneous coverage probabilities for two-sided nominal 0.95 and 0.9 confidence intervals. A balanced experimental design with $J = 3$, $T = 4$, $H = 24$ was simulated according to (7.5). Due to high computational intensity, only few parameter settings with $\mu_i \in [5, 50]$, $\sigma_u = 0.5, \sigma_e = 0.01$ and $\sigma_u = 1, \sigma_e = 1$ were studied.

7.5.2 Summary of results

The simultaneous coverage probability is close to or markedly lower than the nominal level, for both ratio and difference of means. The probability to exclude the true parameter is about equally distributed to lower and upper limits. Most pronounced violations are observed for high abundances in presence of high overdispersion (Table 7.9).

7.5.3 BUGS code and update parameters

```
model{
for(i.obs in 1:n.obs)
{
 epsi[i.obs] ~ dnorm(0, tau.e)
 eta[i.obs] <- inprod(beta[], X[i.obs,])
  + inprod(gamma[],S[i.obs,])
  + epsi[i.obs]
 log(mu[i.obs]) <- eta[i.obs]
 Y[i.obs] ~ dpois(mu[i.obs])
```

```
}
for(i.treat in 1:n.treat)
{
beta[i.treat] ~ dnorm(0, 0.001)
}
for(i.sub in 1:n.subj)
{
gamma[i.sub] ~ dnorm(0,tau.subj)
}
tau.e ~ dgamma(0.001,0.001)
sigma.e <- 1/tau.e
tau.subj ~ dgamma(0.001, 0.001)
sigma.subj <- 1/tau.subj
}
```

Note, that for running this model in `OpenBugs`, the vector `Y`, the structure `X` with dimension `n.obs, n.treat`, the structure `S`, with dimension `n.obs, n.subj` and the numerics `n.obs`, `n.treat`, and `n.subj` have to be provided as input data. Initial values need to provided for the terms `beta[]`, `gamma[]`, `tau.e` , `tau.subj`.

7.5.4 Detailed results

The simulations are based on an MCMC chain of 15000 updates with 5000 first values discarded and 9 out of 10 values discarded in the remaining part of the chain, hence on a sample of $K = 1000$ values from the joint posterior distribution. The estimated coverage probabilities are based on S=1000 simulation runs.

Simulations were run for a model with $H = 24$ independent experimental units, assigned to $J = 3$ treatments, with 8 experimental units assigned to each treatment. Within each treatment, $T = 4$ repeated measures were simulated, resulting in a parameter $\boldsymbol{\beta}$ with $I = 12$ elements β_i. For the fixed quantities of $\mu_i = \exp(\eta_i)$ in Table 7.6, simultaneous coverage probabilities were simulated based on S=1000

draws from the model in Equation (7.5). Table 7.7 shows the contrast coefficients

Table 7.6: Fixed expected values $\exp(\beta_1), ..., \exp(\beta_{12})$ for the 3 treatments and 4 time points in the simulations. In setting 50b, there is a distinct time effect and a time-treatment interaction in treatment 1 compared to treatment 2 and 3, in setting 10b there is a time and a treatment effect but no interaction on the log-scale.

Treatment i	1				2				3			
Time t	1	2	3	4	1	2	3	4	1	2	3	4
50a	50	50	50	50	50	50	50	50	50	50	50	50
50b	50	27	8	10	50	30	10	50	50	30	10	50
10a	10	10	10	10	10	10	10	10	10	10	10	10
10b	10	5	5	10	8	4	4	8	10	5	5	10
5a	5	5	5	5	5	5	5	5	5	5	5	5

c_{mi} used in the simulation. Parameter of interest are the $M = 6$ values $\theta_m = \exp\left(\sum_{i=1}^{I} c_{mi}\beta_i\right)$. For a situation with low overdispersion at the level of individual

Table 7.7: Contrast coefficients applied in the simulation study. In practical application, these contrasts allow to assess whether the dissimilarity in mean abundance to the reference time point 1 is changed in the treatment 1 compared to two reference treatments 2 and 3.

Treatment j	1				2				3			
Time t	1	2	3	4	1	2	3	4	1	2	3	4
i	1	2	3	4	5	6	7	8	9	10	11	12
$m = 1$	1	−1	0	0	−1	1	0	0	0	0	0	0
$m = 2$	1	0	−1	0	−1	0	1	0	0	0	0	0
$m = 3$	1	0	0	−1	−1	0	0	1	0	0	0	0
$m = 4$	1	−1	0	0	0	0	0	0	−1	1	0	0
$m = 5$	1	0	−1	0	0	0	0	0	−1	0	1	0
$m = 6$	1	0	0	−1	0	0	0	0	−1	0	0	1

observations ($\sigma_e = 0.01$) and larger variation at the level of experimental units ($\sigma_u = 0.5$) simultaneous coverage probabilities of 0.95 SCI are shown in Table 7.8. For a situation with about equally large overdispersion at the level of individual observations ($\sigma_e = 1$) and at the level of experimental units ($\sigma_u = 1$), simultaneous coverage probabilities of 0.9 SCI are shown in Table 7.9.

Table 7.8: Simultaneous coverage probability of two-sided nominal 0.95 confidence intervals estimated for ratios $\boldsymbol{\rho}$ and differences $\boldsymbol{\delta}$ from the model in Equation (7.5), assuming the fixed means $\exp(\beta_i), i = 1, ..., 12$ shown in Table 7.6, $\sigma_u = 0.5$, $\sigma_e = 0.1$.

	Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
Setting	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50a	0.941	0.977	0.964	0.951	0.977	0.974
50b	0.944	0.974	0.970	0.959	0.977	0.982
10a	0.931	0.968	0.963	0.937	0.968	0.969
10b	0.947	0.977	0.969	0.952	0.970	0.982
5a	0.960	0.975	0.985	0.965	0.979	0.986

Table 7.9: Simultaneous coverage probability of two-sided nominal 0.9 confidence intervals for ratios $\boldsymbol{\rho}$ and differences $\boldsymbol{\delta}$ from the model in Equation (7.5), assuming the fixed means $\exp(\beta_i), i = 1, ..., 12$ in Table 7.6, $\sigma_u = 1$, $\sigma_e = 1$.

	Ratio $\boldsymbol{\rho}$			Difference $\boldsymbol{\delta}$		
Setting	SCPts	SCPl	SCPu	SCPts	SCPl	SCPu
50a	0.847	0.921	0.926	0.866	0.931	0.935
50b	0.859	0.930	0.929	0.853	0.916	0.937
10a	0.911	0.958	0.953	0.912	0.955	0.957
10b	0.894	0.952	0.942	0.903	0.954	0.949
5a	0.889	0.941	0.948	0.913	0.951	0.962

Chapter 8

Application to real data sets

8.1 Abundance of *Cecidomyiidae* in a genetically modified and three standard crops

In a field trial (Prescher, 2005, personal communication) four varieties were assigned to 32 experimental units, with eight replications each, in a randomized complete block design. One variety, Novum, was a genetically modified variety, a second, Standard, is its near isogenic counter part. The two remaining varieties are conventional varieties, A and B, included in the trial to include the variety specific variability of the abundance of non-target species. At six time points, 12.07., 26.07., 09.08, 24.08, 06.09, and 25.09, the abundance (expressed as number of individuals) of gall midges (family *Cecidomyiidae*) was assessed with an eklektor trap placed in each experimental unit. For modeling the data, the Poisson distribution is assumed for the counts. Since interest might be in variety main effects as well as variety-time interaction, the primary parameters of interest are the 24 means of the four varieties ($j = 1, ..., 4$) at each time point ($t = 1, ..., 6$), modeled on the log-scale. In the design matrix $\boldsymbol{X}$, the means appear in the sequence *Novum,1*, *Novum,2*, ..., *Novum,6*, *Standard,t*, *A,t*, *B,t*. A possibly present correlation between observations from the same experimental units is modeled via the plot effects (corresponding to a compound-symmetry assumption) for which the normal distribution on the log-link

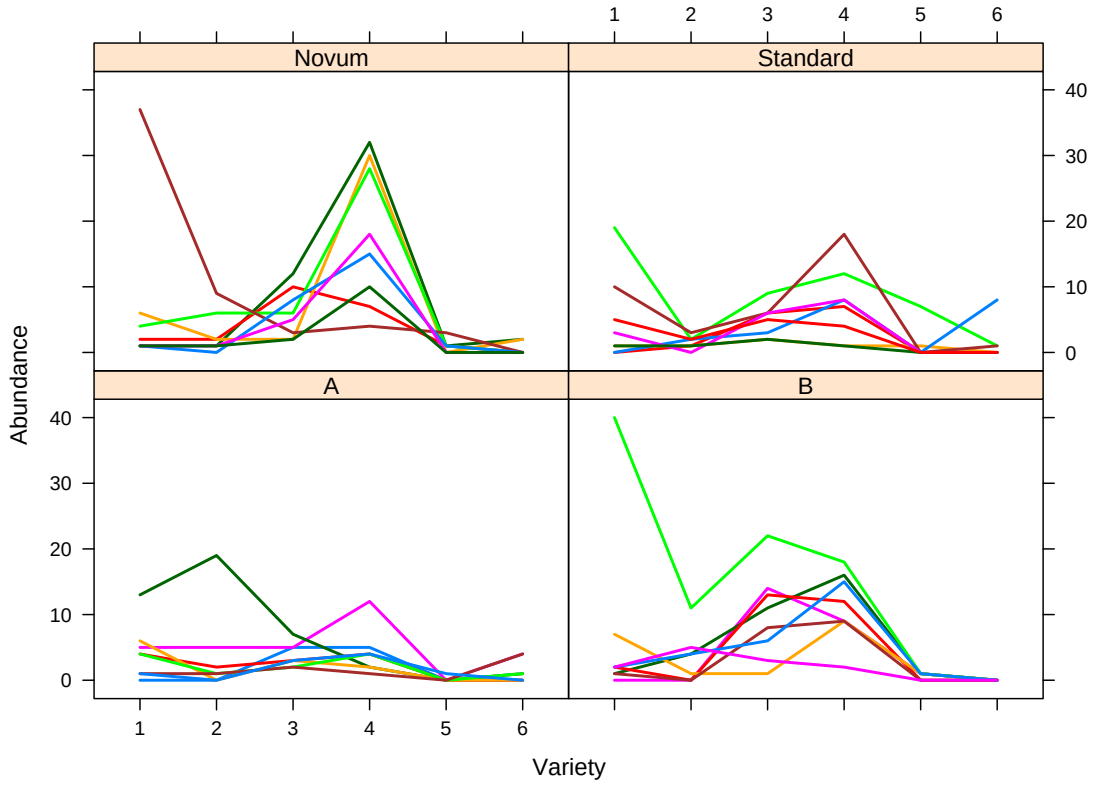

Figure 8.1: Abundance of *Cecidomyiidae* vs. the six time points, separately for the four varieties. Lines join observations from the same experimental units.

is assumed. To account for extra Poisson variability which is neither explained by mean differences nor by the plot effect, an additional normal distributed error term is introduced on the log-scale for individual observations. This model is formally defined in Equation (7.5) in Section 7.5. This model accounts for possible correlations between repeated observations from the same plots and the variability among plots in the field. The model ignores that the treatments are randomized only within blocks. A model that also includes a random block effect and a random effect for block-time interaction would be more appropriate for this dataset, but is beyond the scope of the models investigated in Chapter 7.

Although gall midges feed on plants they are non-target organisms in this context and it is assumed that safety is compromised, when their abundance is decreased in the novel variety. Assume further, that primary interest is in comparison of the varieties pooled over time. On the scale of mean abundance, the hypotheses of interest might be defined as:

$$H_0 : \frac{\mu_{Novum}}{\mu_{Standard}} \leq \rho \cap \frac{\mu_{Novum}}{\mu_A} \leq \rho \cap \frac{\mu_{Novum}}{\mu_B} \leq \rho, \rho < 1 \tag{8.1}$$

$$H_1 : \frac{\mu_{Novum}}{\mu_{Standard}} > \rho \cup \frac{\mu_{Novum}}{\mu_A} > \rho \cup \frac{\mu_{Novum}}{\mu_B} > \rho, \rho < 1, \tag{8.2}$$

where μ_j, $j = 1, ..., 4$ are the expectations of abundance counts on the original scale for each of the four treatments (pooled over time by building geometric means). Simultaneous lower 0.95-confidence limits for the three ratios can be used as tool for rejecting the null-hypothesis with error probability $\alpha = 0.05$, when there is consensus with respect to ρ. The contrast matrix applied to $\boldsymbol{\beta}$ in order to obtain comparisons to control pooled over time then can be conveniently defined by using the Kronecker product $\otimes$:

$$\boldsymbol{C}^{(pool)} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix} \tag{8.3}$$

The parameter of interest is then a column vector of length $M = 3$: $\boldsymbol{\theta} = \exp\left(\boldsymbol{C}^{(pool)}\boldsymbol{\beta}\right)$.

8.1.1 Analysis with non-informative priors

Assume, that there is no prior information available on the parameter of primary interest, $\boldsymbol{\beta}$. For the elements $\boldsymbol{\beta}$ as defined in (7.5), the prior

$$\beta_i \sim N(0, 1000), i = 1, ..., 24 \tag{8.4}$$

is imposed, resulting in an about equal weight contributed to the posterior for $\beta_i \in [0.001, 10000]$. Hence, the posterior is merely conditional to the data. Running the

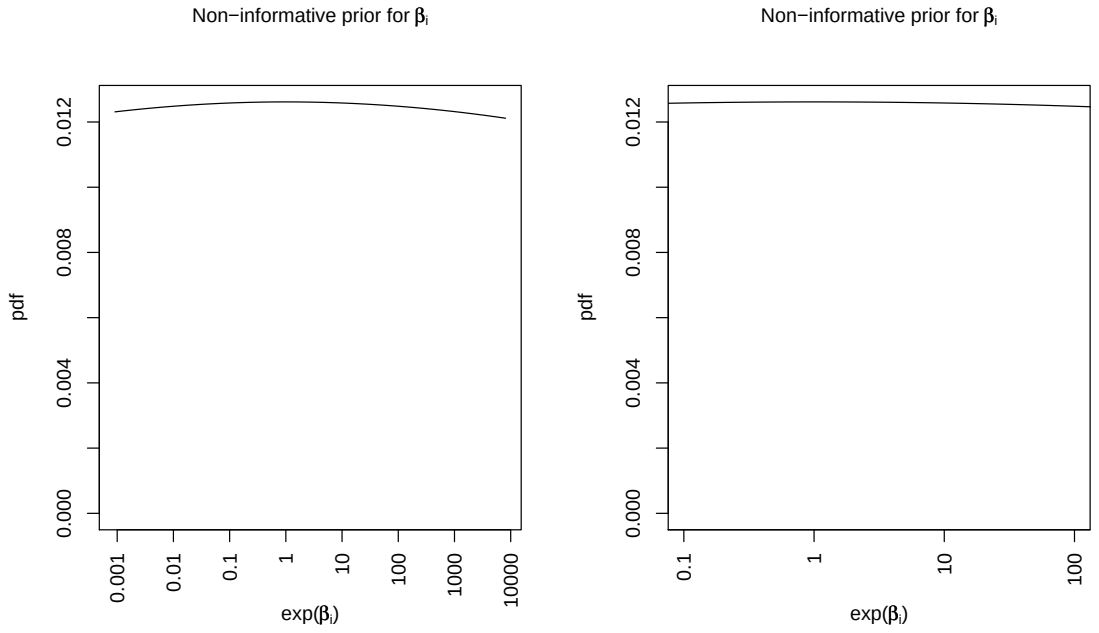

Figure 8.2: Non-informative prior for β_i assumed in the MCMC run leading to the confidence intervals for ratios of mean abundances in Table 8.1. Shown is the pdf of the prior for values $\exp(\beta_i) \in [0.001, 10000]$ and $\exp(\beta_i) \in [0.1, 100]$

model in (7.5.3) with 20000 updates, discarding the first 10000 values from the chain, and retaining one out of ten values in the remaining part of the chain, results in a sample of $K = 1000$ values from the joint posterior. Histograms of the MCMC samples of the primary parameters are shown in Figure 8.3.

Exceptional histograms are those for β_{24} and $\tau_h = 1/\sigma_h^2$ (`tau.subj` in the `BUGS` model). The parameter β_{24} models the mean abundance at the last time point for variety B, where no individual was observed in any of the eight experimental units.

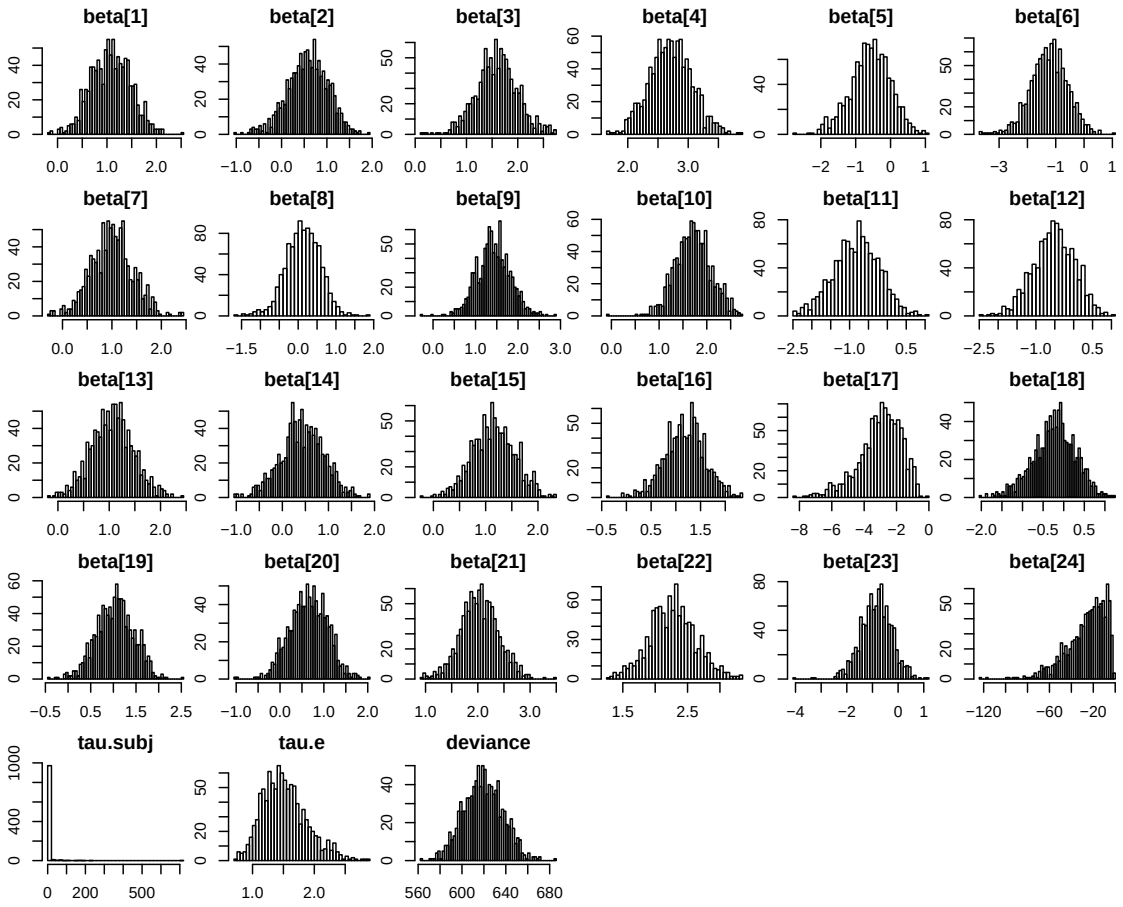

Figure 8.3: Histograms of the marginal posterior distributions from the primary output of the MCMC run, based on $K = 1000$ values, shown are $\beta_1, ..., \beta_{24}$ (`beta[]`), $\tau_e = 1/\sigma_e^2$, (`tau.e`), $\tau_h = 1/\sigma_h^2$ (`tau.subj`) and the deviance.

Being only conditional to the sample, the posterior yields very low values. The posterior for τ_h (`tau.subj`, with σ_h^2 modeling the variance of repeated observations with the same plot) is highly skewed, ranging between 0.9 and 188, with a proportion of 0.95 of the values between 1.7 and 19, and median at 4. The posterior means

Table 8.1: Estimates and simultaneous lower 0.95 confidence limits for the ratio of mean abundance defined in (8.1), (8.2), based on the non-informative prior for the means on the log-scale, defined in (8.4).

Ratio	Estimate	Lower 0.95 limit
Novum/Standard	1.22	0.56
Novum/A	1.81	0.77
Novum/B	35.03	1.23

and the simultaneous confidence intervals for the three ratios of interest are shown in Table 8.1. Due to the low values of β_{24}, the dissimilarity between $Novum$ and B might be overestimated.

8.1.2 Analysis with a weakly informative prior

Using the above priors in (8.4) implies that, for all time points and varieties, mean abundance $\exp(\beta_i)$ of 0.001, or 10000 are considered about as probable as mean abundances of 1 or 10. However, assessing species abundance by counting individuals with sample size 8 implies that *a priori* the taxon is expected to be observable (e.g. mean abundance > 0.1) and practically countable (e.g., mean abundance < 1000) in the experimental setup. Assuming mean abundances centered at $\exp(\beta_i) \cong 2.5$, mean abundances of below 0.1 and above 100 to be relatively rare (probability less than 1/6 for each) and mean abundances below 0.01 or above 1000 to be less probable than 3-4%, could be expressed by the weakly informative prior:

$$\beta_i \sim N\left(\mu = 1, \sigma^2 = 10\right) \tag{8.5}$$

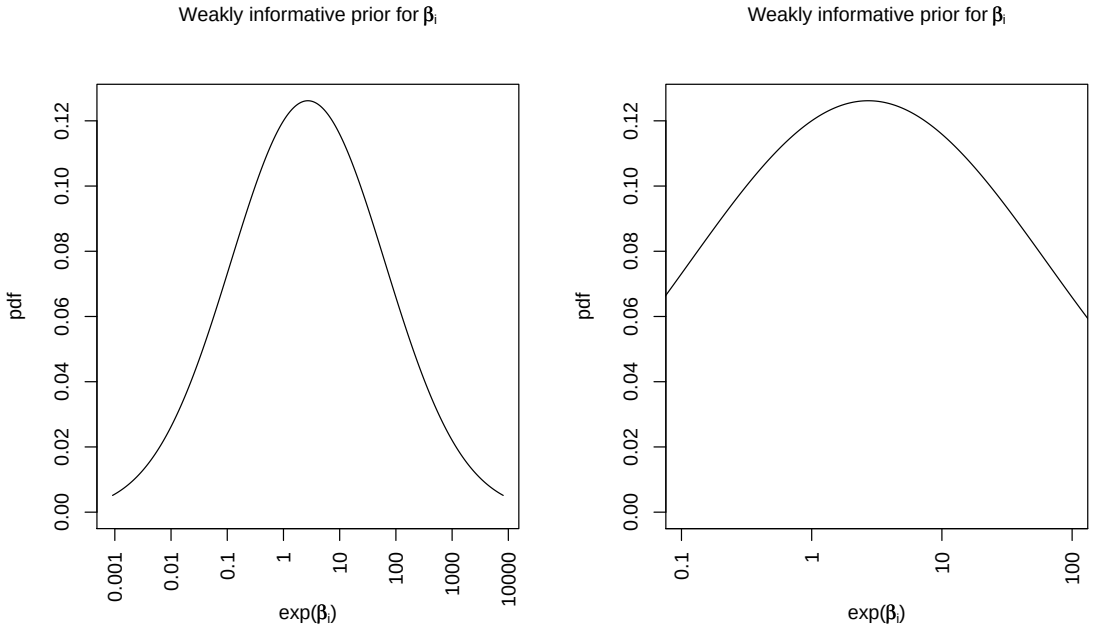

Figure 8.4: Weakly informative prior for β_i assumed in the MCMC run leading to the intervals for ratios of mean abundances in Table 8.2. Shown is the pdf of the prior for values $\exp(\beta_i) \in [0.001, 10000]$ and $\exp(\beta_i) \in [0.1, 100]$

Imposing this weakly informative prior on $\boldsymbol{\beta}$ results in the posterior samples shown in

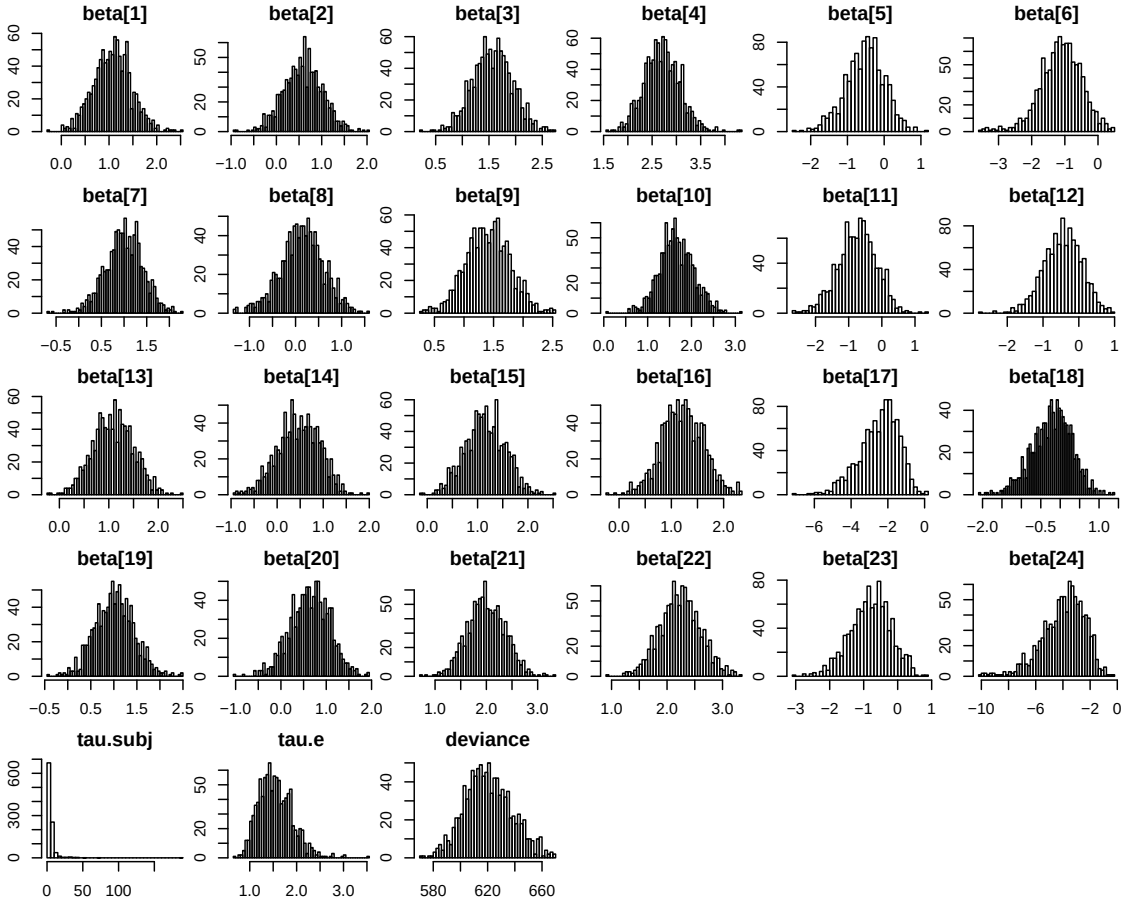

Figure 8.5: Histograms of the marginal posterior distributions from the primary output of the MCMC run with the weakly informative prior in (8.5), based on $K = 1000$ values, shown are $\beta_1, ..., \beta_{24}$ (beta[]), $\tau_e = 1/\sigma_e^2$, (tau.e) $\tau_h = 1/\sigma_h^2$ (tau.subj) and the deviance.

Table 8.2: Estimates and simultaneous lower 0.95 credible limits for the ratio of mean abundance defined in (8.1), (8.2), based on the weakly informative prior for the means on the log-scale defined in (8.5).

	Estimate	lower
Novum/Standard	1.23	0.59
Novum/A	1.65	0.73
Novum/B	1.67	0.72

Figure 8.5 and the simultaneous intervals shown in Table 8.2. Based on the slightly subjective lower bounds for the three ratios of interest and a relevance margin $\rho = 0.5$ one could conclude with 0.95 credibility: The abundance of *Cecidomyiidae* in the GM crop is not relevantly decreased compared to at least one of the conventional varieties, and furthermore, is not relevantly decreased compared to all the conventional varieties, Standard, A, and B.

8.1.3 Exploring interactions

The above decisions are based on the dissimilarity among varieties, pooled over all time points. A secondary experimental question could be to assess the presence of an interaction between time and variety. Interest here is not in all possible interactions in the 4×6 design, but rather in assessing whether and how the ratios of mean abundance in Novum to the three conventional varieties changes in the time points 2, 3, 4, 5, 6 with reference to time point 1:

$$\boldsymbol{C}^{(IA)} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{8.6}$$

The parameter of interest is then the vector $\boldsymbol{\theta} = \exp\left(\boldsymbol{C}^{(IA)}\boldsymbol{\beta}\right)$. The elements of $\boldsymbol{\theta}$ are ratios of two ratios of mean abundances. The first element of $\boldsymbol{\theta}$ relates the ratio of mean abundances in Novum to that in Standard at the first time point to the ratio of mean abundances in Novum to that in Standard at the second time point, $\theta_1 = \frac{\mu_{Novum,Time1}/\mu_{Standard,Time1}}{\mu_{Novum,Time2}/\mu_{Standard,Time2}}$. If this parameter is larger than one, this means that the ratio of mean abundances ($\mu_{Novum}/\mu_{Standard}$) has decreased from time 1 to time 2. If it is smaller than one, the ratio $\mu_{Novum}/\mu_{Standard}$ at time 2 is larger than at time 1. In Figure 8.6, this parameter will be denoted `(N1/S1)/(N2/S2)`. The other standard treatments are denoted `A` and `B`, the time points are identified by their numbers `1,...,6`.

Based on Equation (8.6), Figure 8.6 shows simultaneous 0.95 confidence intervals for the ratios of Novum to the three standard treatment, being compared between time points $t = 2, ..., 6$ and the first time point $t = 1$. The intervals are based on the posterior sample of $K = 1000$ obtained by assuming a non-informative prior with the `BUGS` code in Section 7.5.3. The confidence limits are given in numbers in Table 8.3. Figure 8.7 and Table 8.7 show simultaneous 0.95 credible intervals based on the analysis with the weakly informative prior defined in Equation (8.5) imposed on the elements of $\boldsymbol{\beta}$.

The confidence intervals in Figures 8.6 and 8.7 are very wide, many ranging between 0.1 and 10 fold change in the ratios of mean abundances. Based on these results, it is hard to assume the absence of any change over time in the relevant comparisons to the novel treatment. However, there is also no strong evidence for the presence of such type of an interaction between time and treatment. The only interval which does not contain the value 1 is that for the ratio $\frac{\mu_{Novum,Time1}/\mu_{B,Time1}}{\mu_{Novum,Time6}/\mu_{B,Time6}}$. The data alone indicate that the ratio μ_{Novum}/μ_B is significantly larger at the sixth time point than at the first time point, due to the very low counts in treatment B at time six.

Imposing the prior knowledge leads to a slight shift of the posterior medians of the parameters of interest, and largely increases the upper bound for $\frac{\mu_{Novum,Time1}/\mu_{B,Time1}}{\mu_{Novum,Time6}/\mu_{B,Time6}}$.

Table 8.3: Posterior means and simultaneous 0.95 confidence intervals for the $M = 15$ parameters defined in (8.6) to explore interactions between treatment and time.

Comparison	Estimate	Lower	Upper
(N1/S1)/(N2/S2)	0.71	0.05	6.40
(N1/S1)/(N3/S3)	0.94	0.10	6.70
(N1/S1)/(N4/S4)	0.40	0.04	2.90
(N1/S1)/(N5/S5)	0.88	0.07	14.00
(N1/S1)/(N6/S6)	2.30	0.13	42.00
(N1/A1)/(N2/A2)	0.90	0.11	6.00
(N1/A1)/(N3/A3)	0.66	0.08	4.50
(N1/A1)/(N4/A4)	0.22	0.03	1.20
(N1/A1)/(N5/A5)	0.10	0.00	2.90
(N1/A1)/(N6/A6)	2.80	0.27	40.00
(N1/B1)/(N2/B2)	1.10	0.12	12.00
(N1/B1)/(N3/B3)	1.60	0.21	9.50
(N1/B1)/(N4/B4)	0.68	0.09	3.60
(N1/B1)/(N5/B5)	0.80	0.08	11.00
(N1/B1)/(N6/B6)	0.00	0.00	0.43

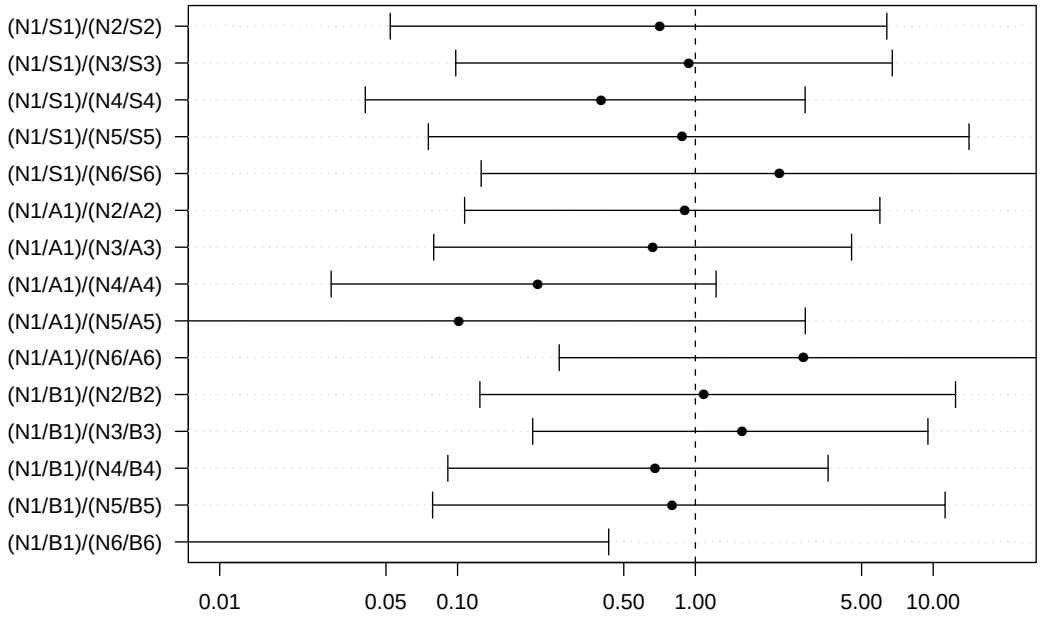

Figure 8.6: Simultaneous 0.95 confidence intervals for the $M = 15$ parameters defined in (8.6) to explore interactions between treatment and time.

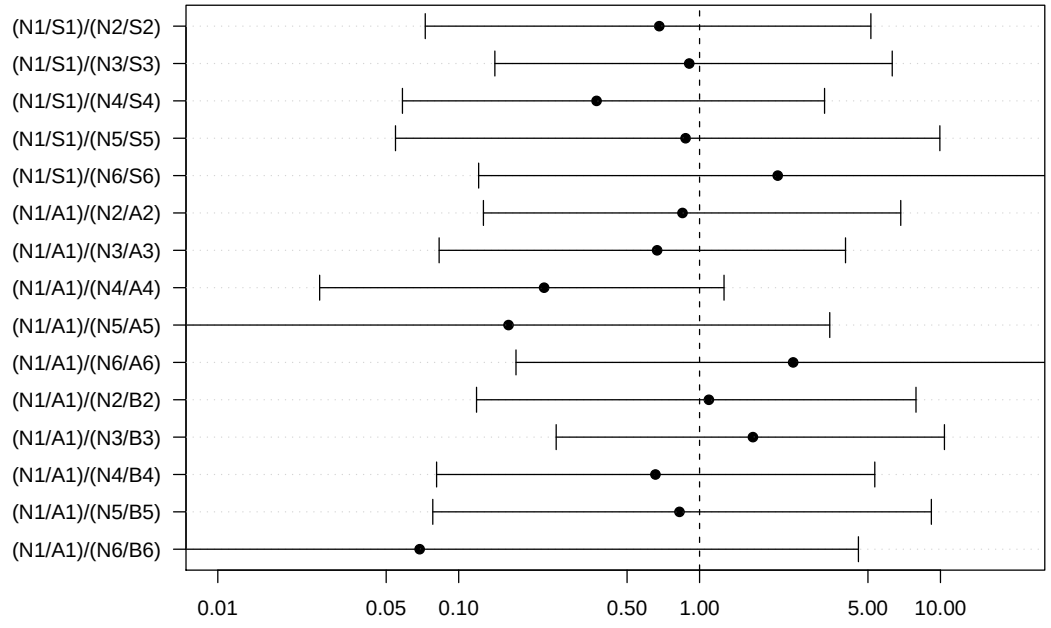

Figure 8.7: Simultaneous 0.95 confidence intervals for the $M = 15$ parameters defined in (8.6) to explore interactions between treatment and time; results based on the analysis with weakly informative priors.

Table 8.4: Posterior means and simultaneous 0.95 confidence intervals for the $M = 15$ parameters defined in (8.6) to explore interactions between treatment and time; results based on the analysis with weakly informative priors.

Comparison	Estimate	Lower	Upper
(N1/S1)/(N2/S2)	0.68	0.07	5.10
(N1/S1)/(N3/S3)	0.91	0.14	6.30
(N1/S1)/(N4/S4)	0.37	0.06	3.30
(N1/S1)/(N5/S5)	0.87	0.06	9.90
(N1/S1)/(N6/S6)	2.10	0.12	38.00
(N1/A1)/(N2/A2)	0.85	0.13	6.80
(N1/A1)/(N3/A3)	0.67	0.08	4.00
(N1/A1)/(N4/A4)	0.23	0.03	1.30
(N1/A1)/(N5/A5)	0.16	0.00	3.50
(N1/A1)/(N6/A6)	2.40	0.17	41.00
(N1/B1)/(N2/B2)	1.10	0.12	7.90
(N1/B1)/(N3/B3)	1.70	0.25	10.00
(N1/B1)/(N4/B4)	0.66	0.08	5.30
(N1/B1)/(N5/B5)	0.83	0.08	9.20
(N1/B1)/(N6/B6)	0.07	0.00	4.60

8.2 Abundance of plant and leaf hoppers in a GM, near isogenic and insecticide treatment

Rauschen et al. (2008) report a field trial where the abundance of plant hoppers and leaf hoppers (suborder *Auchenorrhyncha*) was assessed by visual assessments in a field trial arranged as a randomized complete block design with eight blocks and three treatments. The treatments are a GM-variety (GM), the corresponding near-isogenic line (Iso) and a conventional variety treated with an insecticide (Insecticide). The obtained data are summarized in Figure 8.8: The following model is assumed

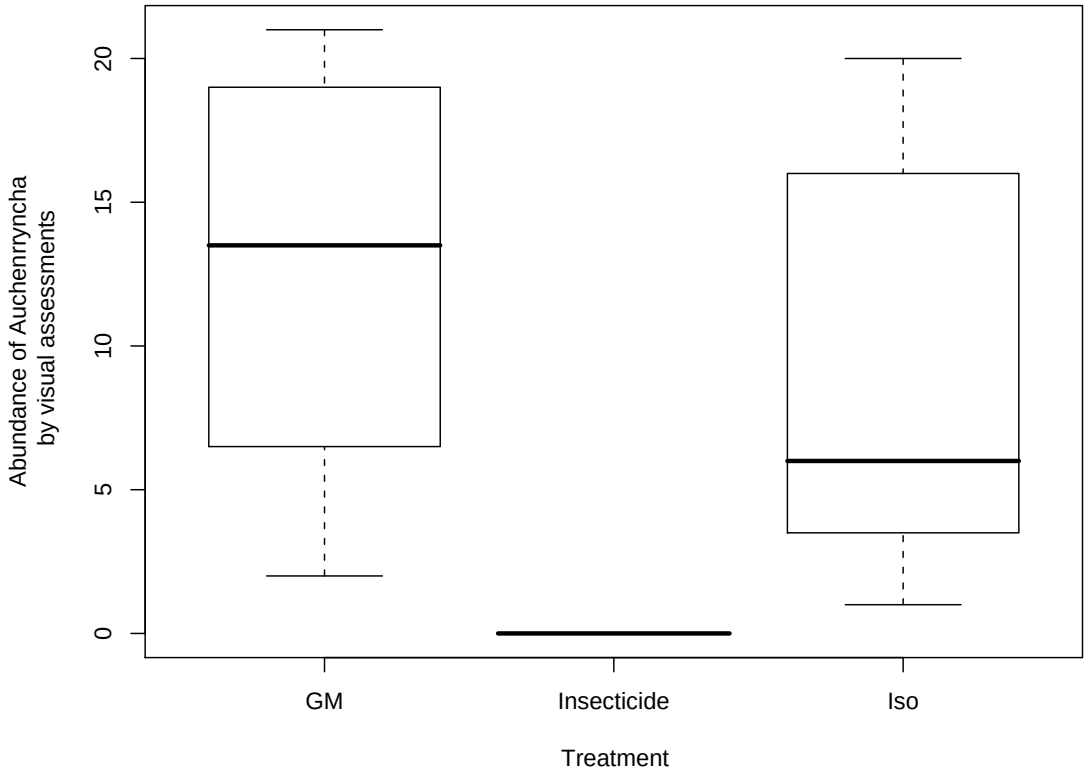

Figure 8.8: Boxplots of the abundance of *Auchenorrhyncha* in the three treatments. The bold lines mark the sample median, the lower and upper end of the boxes mark the 0.25 and 0.75 sample quantiles and the whiskers mark the minimal and maximal values.

for the data:

$$\begin{aligned} Y_n &\sim NB(\mu_n, \tau) \\ \mu_n &= \exp(\eta_n) \\ \eta_n &= \sum_{i=1}^{I} x_{ni}\beta_i + \sum_{j=1}^{J} z_{nj}\gamma_j \\ \gamma_j &\sim N(0, \sigma_u^2) \end{aligned} \tag{8.7}$$

Primary interest is in comparison of the treatments (i.e., the β_i, $i = 1, 2, 3$), whereas the potential effects of the blocks (i.e., the γ_j, $j = 1, ..., 8$) are modeled as normal distributed on the log-link by the hyper parameter σ_u.

In safety assessment, the hypothesis of primary interest then is:

$$H_{01} : \mu_{GM}/\mu_{Iso} \leq \underline{\rho}_1 \cup \mu_{GM}/\mu_{Iso} \geq \overline{\rho}_1 \tag{8.8}$$

$$H_{11} : \mu_{GM}/\mu_{Iso} > \underline{\rho}_1 \cap \mu_{GM}/\mu_{Iso} < \overline{\rho}_1 \tag{8.9}$$

with $\underline{\rho}_1 < 1$ and $\overline{\rho}_1 > 1$ specified *a priori*. A marginal 0.95 lower and 0.95 upper confidence limit for μ_{GM}/μ_{Iso} are adequate statistical tools for deciding on these hypotheses and presenting the results, especially when there is no *a priori* consensus concerning $\underline{\rho}_1$ and $\overline{\rho}_1$.

Of secondary interest might be the comparison of the GM line to the insecticide treatment. Here it might be of interest to show the non-inferiority or superiority of GM vs. Insecticide.

$$H_{02} : \mu_{GM}/\mu_{Insecticide} \leq \underline{\rho}_2 \tag{8.10}$$

$$H_{12} : \mu_{GM}/\mu_{Insecticide} > \underline{\rho}_2 \tag{8.11}$$

where ρ_2 might be chosen $\rho_2 < 1$, $\rho_2 = 1$, or $\rho_2 > 1$, depending on whether it is of interest to show non-inferiority, marginal dissimilarity or superiority of GM compared to Insecticide. A lower 0.95 confidence limit for the ratio $\mu_{GM}/\mu_{Insecticide}$ is an appropriate tool to decide on such hypotheses, independent of the choice of ρ_2. Note, that the hypotheses (8.8), (8.9) and (8.10), (8.11) imply different aspects of safety assessment for the GM variety and hence might be adequate not to control the familywise error for both. Still, if interest is in controlling the familywise error for the two experimental questions, it is reasonable to impose an *a priori* order on the hypotheses, where again marginal confidence limits for the two parameters of interest are appropriate (Hothorn and Lehmacher, 1991).

8.2.1 Analysis with non-informative prior

$$\beta_i \sim N(0, 1000), i = 1, 2, 3 \tag{8.12}$$

Table 8.5: Posterior medians, lower 0.95 and upper 0.95 confidence limits for the parameters of interest in (8.8) and (8.10) when analysed with the non-informative prior for β_i in (8.12).

Comparison	Estimate	Lower 0.95 limit	Upper 0.95 limit
GM/Iso	1.5	0.69	3.3
GM/Insecticide	22.4	8.3	61.4

8.2.2 Analysis with a weakly informative prior

$$\beta_i \sim N(0, 10), i = 1, 2, 3 \tag{8.13}$$

Table 8.6: Posterior medians, lower 0.95 and upper 0.95 confidence limits for the parameters of interest in (8.8) and (8.10) when analysed with the weakly informative prior for β_i in (8.13).

Comparison	Estimate	Lower 0.95 limit	Upper 0.95 limit
GM/Iso	1.6	0.68	3.3
GM/Insecticide	23.4	8.8	65.8

Assuming relevance margins of $\underline{\rho}_1 = 0.5$ and $\overline{\rho}_1 = 2.0$ for the equivalence problem of primary interest, one can not conclude for equivalence of GM compared to Iso, since an increase of abundance in GM to more than 3.3 times the abundance in Iso can not be ruled out with high confidence. However, one may conclude with 0.95 confidence that the abundance of plant and leaf hoppers is relevantly increased in GM compared to their abundance in the insecticide treatment. Based on the non-informative prior,

one might conclude with 95% confidence that the abundance of *Auchenorrhyncha* in the GM variety is at least 8.3 times higher than in the conventional variety with insecticide treatment. However, based on the simulation study in Chapter 5, one can know that bounds constructed in this way exclude the true value more often than in 5% of the cases. Hence, the actual frequentist confidence in the above statement is somewhat lower than 95%, even if the assumptions of the statistical model hold true.

Comparing the results of the analysis with a non-informative and a weakly informative prior shows, that the confidence limits of interest are not very sensitive for the particular choices of prior distributions.

Chapter 9

Discussion

The problem of local and simultaneous confidence intervals for dissimilarity in mean abundance among two or several samples is a relevant problem with applications in monitoring the ecological impact of novel practices, chemicals or varieties on non-target species. Similar applications with increasing public interest arise in fishery management and epidemiology of parasites. The construction of valid confidence intervals and simultaneous confidence sets for ratios or differences of means abundances has not gained much attention in the literature so far. Especially the problem of simultaneous confidence intervals has not been considered in detail. This work briefly reviews methods to construct marginal and simultaneous confidence intervals for dissimilarity in mean abundance. After briefly comparing their performance in a relevant extreme data realization, the focus is on the frequentist coverage probability of marginal and simultaneous confidence intervals constructed using the MCMC. It is shown that, in principle, the application of Bayesian models with non-informative prior distributions in the Gibbs sampler, may yield correct frequentist marginal and simultaneous confidence intervals. The frequentist validity of such confidence intervals is shown for a variety of statistical models, including the completely randomized one-way layout and some simple hierarchical models with clustered experimental units or repeated measurements. However, the simulation studies show that for small sample sizes and low abundances, both marginal and simultaneous confidence intervals derived from the Gibbs sampler perform liberal. I.e., in the frequentist way

of thinking they cover an assumed true parameter value less often than prespecified. Hence, when applied in safety assessment as a tool for deciding in a proof of equivalence or non-inferiority, the prespecified consumers' risk is inflated. Especially, the application of the discussed methods can not be recommended when abundances are very small and when the species of interest is absent in at least one group. Under favorable conditions, i.e., with mean abundances at least 10 and groupwise sample sizes at least 20, both, marginal and simultaneous confidence intervals have a coverage probability close to the nominal level. Additionally, the simulation studies show that it is surprisingly simple to construct valid (simultaneous) confidence intervals for differences of mean abundances, which is a non-canonical measure of dissimilarity in terms of generalized linear models with log-link. However, its use in safety assessment is limited.

Using the Gibbs sampler to derive simultaneous confidence intervals, as discussed in this work, has a number of drawbacks. First, the communication between two software packages, `R` and `OpenBUGS`, complicates the technical execution of the statistical analysis. Further, without much theoretical background, the definition of many different models is possible and MCMC will produce an output as long as the BUGS code is syntactically correct and fits to the input data. Hence, in practical application, where computation time is not as problematic as in simulation studies, more effort should be put on thoroughly assessing the convergence, e.g. by running the model with several chains, initialized by values scattered over the parameter space (e.g. Gelman et al., 2004; Browne and Draper, 2005). Also, it is recommended to chose the number of updates and the number of values discarded at the beginning of the chains much larger, than in the simulation studies here. Still the resulting confidence intervals should be interpreted with care, since an at least slightly liberal performance is the common result of all simulation studies in this work.

The results of this work can not be more than an initial proof of concept for deriving frequentist (simultaneous) confidence intervals for difficult distributional assumptions from MCMC. The reasons are manifold: The models considered have been simulated for only a very limited number of experimental designs. Especially

for the hierarchical models, also more extreme designs concerning the number of clusters, replications within clusters, etc. should be considered. Further, focus was merely on factorial arrangements and subsequent multiple treatment comparisons, a problem which is rarely discussed in the Bayesian literature and user manuals. In real life, problems often involve combinations of factorial arrangements and numerical covariates; the problems arising in such settings have not been considered here. Moreover, the hierarchical models considered in this work are limited to rather simple structures. In the problem of clustered experimental units a possible interaction of treatments and environments has not been included in the considered models. Moreover, randomization structures with more than one level of nesting are frequently observed in practice. For such models, the design matrices accounting for the randomization structure will be more complex and several hyper parameters have to be specified for the different random effects. When considering designs with repeated measurements within experimental units, other than the compound symmetry assumption considered here might be reasonable. Since the multivariate normal distribution is implemented in `OpenBUGS`, assumptions on the variance-covariance structure of observations might be included in models for repeated measures. However, this introduces the problem of defining prior distributions for the parameters in variance-covariance structures. Hints for defining more complex hierarchical models can, e.g., be found in Browne and Draper (2005); Gelman (2006); Kass and Natarajan (2006); Zhao et al. (2006).

Another limitation of the methods discussed here are the distributional assumptions. Being based on the likelihoods following from the probabilistic model assumed for the dependence of the observations on the parameters and hyper parameters, the Gibbs sampler is a 'parametric' method. Hence, the inferential results rely on these assumptions. However, in agricultural and ecological studies, sample size is usually too small to thoroughly assess the validity of the distributional assumptions. Following this uncertainty, the problem of model selection arises. Due to the simplicity by which different models with different distributional assumptions can be defined in BUGS, several slightly different models for the same data situations are possible.

For example, in the above models, the means are modeled by normal distributions on the log-link. Alternatively, they could be modeled by gamma distributions on the original positive scale, as is done in the problem of HGLMs (Lee et al., 2006). Similarly, in some of the above models, overdispersion has been modeled by the negative binomial distribution (corresponding to a Poisson mixture with gamma-distributed expectations) or by imposing a normal distribution on the residuals on the log-scale. In this work, the problem how to decide among several models in practice has not been considered. Hence, for completion of a recommendation of a statistical method, its robustness in case of violated assumptions has to be assessed, which has not been done in this work. Finally, a related problem is the choice of the link function: here, the log-function has been chosen merely by convention and convenience, what is not necessarily the best option in practice.

Chapter 10

Extensions and Outlook

The statistical procedures discussed above may be straightforwardly extended for application in situations with similar experimental designs but different distributional assumptions. Most prominent example is that of constructing confidence intervals for parameters describing the dissimilarity among binomial proportions. For the problem of non-canonical parameters, as are the risk difference and the risk ratio, even in the simple two-sample case a number of methods has been proposed and is still under discussion (Newcombe, 1998; Agresti and Caffo, 2000; Brown and Li, 2005, to name a few references). Recently, the distribution for the difference of two proportions has been published (Nadarajah and Kotz, 2007), however, without making a confidence interval method available based on these findings. Only some approximate solutions are available for multiple comparisons among proportions: while in some approaches only special cases are considered (Holford, 1989; Piegorsch, 1991; Agresti et al., 2008; Schaarschmidt et al., accepted a), Bretz and Hothorn (2002) and Schaarschmidt et al. (accepted b) provide a solution for a general contrasts (on the scale of the proportion) with focus on large sample hypotheses testing, and approximate small sample confidence intervals, respectively. However, instead of deriving exact distributions (Nadarajah and Kotz, 2007) or using the multivariate normal approximation (Bretz and Hothorn, 2002; Schaarschmidt et al., accepted b), the methods discussed in this work might be used to sample from the distribution of interest for user-defined contrasts of proportions. The odds ratio,

risk ratio or risk difference might be defined as parameters of interest. An approach via the Gibbs sampler is especially appealing because, in some applications as the assessment of carcinogenicity and toxicity of compounds, prior information on the binomial parameter is available at least for untreated control groups (Tarone, 1982; Tamura and Young, 1986). However, when interest is in the small sample properties of such methods, these should be assessed for relevant parameter settings in another simulation study.

Moreover, the method of Besag et al. (1995) to construct confidence sets based on samples of the joint distribution of the parameters of interest might be applied to the output of other algorithms to draw such samples in the frequentist context. Such algorithms could be parametric or non-parametric bootstrap stratified for multiple treatment comparisons (e.g. Davison and Hinkley, 1997), where the percentile method of Besag et al. (1995) could be a competitor to the simultaneous confidence intervals proposed by Beran (1990); Mandel and Betensky (2008).

A relevant problem related to safety assessment for non-target species is to construct simultaneous confidence intervals for biodiversity indices. The motivation is simple: the considerations of this work and the power assessment of Gerhard and Schaarschmidt (2007) show that for the safety assessment concerning rare species, available statistical methods do not perform acceptably and, if still applied, do not have sufficient power in a proof of safety with commonly accepted safety margins. Straightforward ways out are the inclusion of prior information (as suggested by Dixon et al., 2005, for a related problem), focusing on species with sufficient abundance (e.g. Rauschen et al., 2008) or the summary of species according to ecological criteria (feeding, behavior, exposure to the potentially hazardous agent). However, additionally to showing that the abundance of dominant species and important ecological groups is not severely affected, it could be of interest to show that the multitude of rare species is not severely compromised. One way to tackle this problem are diversity indices, for examples according to Shannon or Simpson (Magurran, 1988). The construction of marginal and simultaneous confidence intervals for differences of these indices has been considered already (Fritsch and Hsu, 1999; Rogers and

Hsu, 2001), though the proposed methods can not account for the overdispersion usually observed in ecological field data. Since an analytical approach for overdispersed multinomial data can be expected to be difficult, a bootstrap in combination with the approach of Besag et al. (1995) may be a viable option that includes major sources of uncertainty.

Alternatively, the method of generalized pivotal quantities (Weerahandi, 1993) might be used to obtain the joint distribution of parameter estimates in settings with non-trivial distributional assumptions. For example, one could attempt to use the methods described in Chapter 4 for extending the methods of Chen and Zhou (2006) for ratio and difference of log normal means to the problem of multiple comparisons in the one-way layout. However, compared to these approaches, the Gibbs sampler appears to be more convenient and flexible when hierarchical models are of interest.

Bibliography

Agresti, A. (2003). Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research* **12**, 3-21.

Agresti, A., Bini, M. , Bertaccini, B. and Ryu, E. (2008). Simultaneous confidence intervals for comparing binomial parameters. *Biometrics* Early View, February 2008.

Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician* **54**, 280-288.

Agresti, A. and Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics* **61**, 515-523.

Andow, D.A. (2003). Negative and positive data, statistical power, and confidence intervals. *Environmental Biosafety Research* **2**, 1-6.

Anscombe, F.J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* **5**, 165-173.

Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* **37**, 358-382.

Beninca, E., Huisman, J., Heerklos, R., Jöhnk, K.D., Branco, P., Van Nes, E.H., Scheffer, M., Ellner, S.P. (2008). Chaos in a long-term experiment with a plankton community. *Nature* **451**, 822-825.

Beran, R. (1990). Refining bootstrap simultaneous confidence intervals. *Journal of the American Statistical Association* **85**, 417-426.

Berberich, S.A., Ream, J.E., Jackson, T.L., Wood, R., Stipanovic, R., Harvey, P., Patzer, S. and Fuchs, R.L. (1996). The composition of insect-protected cottonseed is equivalent to that of conventional cottonseed. *Journal of Agricultural and Food Chemistry* **44**, 365-371.

Berger, R.L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295-300.

Berger, R.L. and Hsu, J.C. (1996). Bioequivalence trials, intersection union tests and equivalence confidence sets. *Statistical Science* **11**, 283-319.

Berger, J.O. and Sun, D. (2008). Objective priors for the bivariate normal model. *The Annals of Statistics* **36**, 963-982.

Berry, D.A. and Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference* **82**, 215-227.

Besag, J., Green, P., Higdon, D., Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science* **10**, 3-66.

Bliss C.I. and Fisher, R.A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics* **9**, 176-200.

Bofinger E. and Bofinger M. (1993). Equivalence of normal means compared with a control. *Communications in Statistics - Theory and Methods* **22**, 3117-3141.

Bofinger E. and Bofinger M. (1995). Equivalence with respect to control - stepwise tests. *Journal of the Royal Statistical Society Series B* **57**, 721-733.

Breslow, N. (1990). Tests of hypotheses in overdispersed poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* **85**, 565-571.

Bretz, F. (1999). Powerful modifications of Williams test on trend. Dissertation, Universität Hannover.

Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis* **50**, 1735-1748.

Bretz, F. and Hothorn, L. (2002). Detecting dose-response using contrasts: asymptotic power and sample size determination for binomial data. *Statistics in Medicine* **21**, 3325-3335.

Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science* **16**, 101-133.

Brown, L.D. and Li, X. (2005). Confidence intervals for two sample binomial distributions. *Journal of Statistical Planning and Inference* **130**, 359-375.

Browne, W.J. and Draper, D. (2005). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* **1**, 437-524.

Campbell, N.L., Young, L.J., and Capuano, G.A. (1999). Analyzing over-dispersed count data in two-way cross-classification problems using generalized linear models. *Journal of Statistical Computation and Simulation* **63**, 263-281.

Carlin, B.P., Clark, J.S. and Gelfand, A.E. (2006). Elements of hierarchical Bayesian inference. In: Clark, J.S, and Gelfand, A.E. (Eds.): *Hierarchical Modelling for the Environmental Sciences*. Oxford University Press, Oxford.

Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Second edition. Duxbury, Pacific Grove.

Chen, J.-S. and Jennrich, R.I.(1996): The signed root deviance profile and confidence intervals in likelihood analysis. *Journal of the American Statistical Association* **91**, 993-998.

Chen, J. and Sarkar, S.K. (2004). Multiple testing of response rates with a control: a Bayesian stepwise approach. *Journal of Statistical Planning and Inference* **125**, 3-16.

Chen, Y.-H. and Zhou, X.-H. (2006). Interval estimates for the ratio and the difference of two lognormal means. *Statistics in Medicine* **25**, 4099-4113.

Clark, S.Z., Rothery, P. and Perry, J.N. (2006). Farm evaluations of spring-sown genetically modified herbicide-tolerant crops: a statistical assessment. *Proceedings of the Royal Society B* **273**, 237-243.

Clayton, R.G. (1996). Generalized linear mixed models. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (Editors): *Markov Chain Monte Carlo in Practice.* Chapman and Hall, London.

Congdon, P. (2006). *Bayesian Statistical Modelling.* Second Edition. John Wiley & Sons, Ltd. Chichester.

Datta, G.S. (1996). On priors providing frequentist validity of Bayesian inference for multiple parametric functions. *Biometrika* **83**, 287-298.

Datta, G.S. and Ghosh, J.K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika* **82**, 37-45.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

Dilba, G. (2005). *Simultaneous inference for ratios of location parameters.* PhD thesis, Naturwissenschaftliche Fakultät, Universität Hannover.

Dilba, G. (2006). Simultaneous credible intervals for ratios of means. Oral contribution to the conference: Evaluation im Gesundheitswesen, Bochum.

Dilba, G., Bretz, F., and Guiard, V. (2006). Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal of Statistical Planning and Inference* **136**, 2640-2658.

Dixon, P.M., Ellison, A.M., Gotelli, N.J. (2005). Improving the precision of estimates of the frequency of rare events. *Ecology* **86**, 1114-1123.

Duncan, D.B.(1965). A Bayesian approach to multiple comparisons. *Technometrics* **7**, 171-222.

Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096-1121.

Dunnett, C.W. and Tamhane, A.C. (1997). Multiple testing to establish superiority/equivalence or a new treatment compared with k standard treatments. *Statistics in Medicine* **16**, 2489-2506.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association* **99**, 96-104.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association* **102**, 93-103.

Efron, B., Tibshirani, R., Storey, J.D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* **96**, 1151-1160.

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecology Letters* **7**, 509-520.

Fletcher D., MacKenzie D., Villouta E. (2005). Modelling skewed data with many zeros: A simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics* **12**, 45-54.

Fritsch, K.S., and Hsu, J.C. (1999). Multiple comparison of entropies with application to Dinosaur biodiversity. *Biometrics* **55**, 1300-1305.

Gelfand, A.E. Sahu, S.K. and Carlin, B.P. (1995). Efficient parametrisations for normal linear models. *Biometrika* **82**, 479-488.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. Second Edition. Chapman and Hall/CRC. Boca Raton.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis* **1**, 515-534.

Gerhard, D. (2008, personal communication). Institut für Biostatistik, Leibniz Universität Hannover.

Gerhard D. and Schaarschmidt F. (2007). Proof of Safety for non-target species: a confidence interval based approach. In Piepho H.-P., Bleiholder H. (Eds.): *Proceedings of the International Symposium Agricultural Field Trials - Today and Tomorrow*, Verlag Breuer, Stuttgart.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In Bernardo, J.M., Berger, J., Dawid, A.P. and Smith, A.F.M. (Eds.) *Bayesian Statistics 4* Oxford University Press, Oxford.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc., Hoboken.

Ghosh, M. and Kim Y.-H. (2001). The Behrens-Fisher problem revisited: A Bayes-frequentist synthesis. *The Canadian Journal of Statistics* **29**, 5-17.

Ghosh, P. and Rosner, G.L. (2007). A semi-parametric Bayesian approach to average bioequivalence. *Statistics in Medicine* **26**, 1224-1236.

Ghosh, M., Yin, M. and Kim Y.-H. (2003). Objective Bayesian inference for ratios of regression coefficients in linear models. *Statistica Sinica* **13**, 409-422.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). Introducing Markov Chain Monte Carlo. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (Eds.): *Marcov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Hauschke, D. (1999). Biometrische Methoden zur Planung und Auswertung von Sicherheitsstudien. Habilitationsschrift Fachbereich Statistik, Universität Dortmund.

Herman, R.A., Philips, A.M., Collins, R.A., Tagliani, L.A., Claussen, F.A., Graham, C.D., Bickers, B.L., Harris, T.A., Prochaska, L.M. (2004). Compositional Equivalency of Cry1F Corn Event TC6275 and Convetional corn (Zea Mays L.). *Journal of Agricultural and Food Chemistry* **52**, 2726-2734.

Holford, T.R., Walter, S.D. and Dunnett, C.W. (1989). Simultaneous interval estimates of the odds ratio in studies with two or more comparisons. *Journal of*

Clinical Epidemiology **42**, 427-434.

Hothorn, T., Bretz, F. and Genz, A. (2001). On multivariate t and Gauss probabilities in R. *R News* **1 (2)**, 27-29.

Hothorn, L.A. and Lehmacher, W. (1991). A simple testing procedure 'Control versus k treatments' for one-sided ordered alternatives, with application in toxicology. *Biometrical Journal* **33**, 179-189.

Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal* **50**, 346-363.

Hyndman, R.J. and Fan, Y (1996). Sample quantiles in statistical packages. *The American Statistician* **50**, 361-365.

Ives, A.R., Einarsson, A., Jansen, V.A.A., Gardarsson, A. (2008). High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. *Nature* **452**, 84-87.

Ji, Y., Lu, Y. and Mills, G.B. (2008). Bayesian models based on test statistics for multiple hypothesis testing problems. *Bioinformatics* **24**, 943-949.

Johnson, N.L., Kotz, S., Balakrishnan, N.(1994). *Continuous Univariate distributions.* Volume 1. Second edition. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., New York.

Johnson, N.L., Kotz, S., Kemp, A.W.(1993). *Univariate discrete distributions.* Second edition. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., New York.

Kass, R.E. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (Comment on article by Brown and Draper). *Bayesian Analysis* **1**, 535-542.

Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Continuous Multivariate Distributions* Volume 1. Models and Applications. Second edition. John Wiley and Sons, Inc., New York.

Laster, L.L. and Johnson, M.F. (2003). Non-inferirority trials: the 'at least as good as' criterion. *Statistics in Medicine* **22**, 187-200.

Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.

Lee, Y., Nelder, J.A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects. Unified Analysis via H-likelihood.* Chapman and Hall/CRC, Boca Raton.

Lindley, D.V.(1998). Decision analysis and bioequivalence trials. *Statistical Science* **13**, 136-141.

Magurran, A.E. (1988). *Ecological Diversity and its Measurement.* Princeton University Press, Princeton, New Jersey.

Mandel, M. and Betensky, R.A. (2008). Simultaneous confidence intervals based on the percentile bootstrap approach. *Computational Statistics and Data Analysis* **52**, 2158-2165.

McCullagh, P. and Nelder, J.A.(1989). *Generalized Linear Models.* Chapman and Hall, London New York.

McCulloch, C.E. and Searle, S.R. (2001): *Generalized, Linear and Mixed Models.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., NewYork.

Nadarajah, S., Kotz, S. (2007). Statistical distribution of the difference of two proportions. *Statistics in Medicine* **26**, 3518-3523.

Nashimoto, K. and Wright,F.T. (2008). Bayesian multiple comparisons of simply ordered means using priors with a point mass. *Computational Statistics and Data Analysis* **52**, 5143-5153.

Nelson, P.R. (1989). Multiple comparisons of means using simultaneous confidence intervals. *Journal of Quality Technology* **21**, 232-289.

Newcombe, R.G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* **17**, 873-890.

Nicolaou, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *Journal of the Royal Statistical Society B* **55**, 377-390.

Nobile, A. and Green, P.J. (2000). Bayesian analysis of factorial experiments by mixture modelling. *Biometrika* **87**, 15-35.

Obert, J.C., Ridley, W.P., Schneider, R.W., Riodan, S.G., Nemeth, M.A., Trujillo, W.A., Breeze, M.L., Sorbet, R. and Astwood, J.D. (2004). The composition of grain and forage from glyphosate tolerant wheat MON 71800 is equivalent to that of conventional wheat (Triticum aestivum L.). *Journal of Agricultural and Food Chemistry* **52**, 1375-1384.

Peers, H.W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *Journal of the Royal Statistical Society B* **27**, 9-16.

Pennello, G.A. (2007). Duncan's k-ratio Bayes rule approach to multiple comparisons: an overview. *Biometrical Journal* **49**, 78-93.

Perry, J.N., Rothery, P., Clark, S.J., Heard, M.S., and Hawes, C. (2003). Design, analysis and statistical power of the farm-scale evaluations of genetically modified herbicide-tolerant crops. *Journal of Applied Ecology* **40**, 17-31.

Piegorsch, W.W. (1991). Multiple comparisons for analyzing dichotomous response. *Biometrics* **47**, 45-52.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2007). coda: Output analysis and diagnostics for MCMC. R package version 0.12-1.

Potts, J.M. and Elith, J. (2006). Comparing species abundance models. *Ecological Modelling* **199**, 153-163.

Prescher, S. (2005, personal communication). Biologische Bundesanstalt für Land- und Forstwirtschaft, Berlin und Braunschweig.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-

900051-07-0, URL http://www.R-project.org.

Rauschen, S., Eckert, J., Schaarschmidt, F., Schuphan, I., Gathmann, A. An evaluation of methods for assessing the impacts of Bt-maize MON810 cultivation and pyrethroid insecticide use on Auchenorrhyncha (Planthoppers and Leafhoppers). To appear, accepted for publication in *Agricultural and Forest Entomology*.

Reynolds, T.L., Nemeth, M.A., Glenn, K.C., Ridley, W.P. and Astwood, J.D. (2005). Natural variability of metabolites in maize grain: Differences due to genetic background. *Journal of Agricultural and Food Chemistry* **53**, 10061-10067.

Rigby, R.A. and Stasinopoulos D.M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics* **54**, 507-554.

Rogers, J.A. and Hsu, J.C. (2001). Multiple comparisons of biodiversity. *Biometrical Journal* **43**, 617-625.

Saha, K.K., Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics* **61**, 179-185.

Schaarschmidt, F. (2005). *Binomial group testing - Design and Analysis.* Diplomarbeit am Lehrgebiet Biostatistik, Fachbereich Gartenbau, Universitaet Hannover.

Schaarschmidt, F. (2008). pairwiseCI: Confidence intervals for two sample comparisons. R package version 0.1-13.

Schaarschmidt, F., Biesheuvel, E. and Hothorn, L.A. (a) Asymptotic simultaneous confidence intervals for many-to-one comparisons of binary proportions in randomized clinical trials. Accepted for publication in *Journal of Biopharmaceutical Statistics*.

Schaarschmidt, F., Sill, M. and Hothorn, L.A. (b) Approximate simultaneous confidence intervals for multiple contrasts of binomial proportions. Accepted for publication in *Biometrical Journal*.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of*

Pharmacokinetics and Biopharmaceutics, **15**, 657-680.

Scott, J.G. and Berger, J.O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* **136**, 2144-2162.

Selwyn, M.R., Dempster, A.P. and Hall, N.R. (1981). A Bayesian approach to bioequivalence for the 2 x 2 changeover design. *Biometrics* **37**, 11-21.

Shang, J., Cavanaugh, J.E. and Wright, F.T. (2008). A Bayesian multiple comparison procedure for order-restricted mixed models. *International Statistical Review* **76**, 268-284.

Shi L. and Bai P. (2008). Bayesian confidence interval for the difference of two proportions in the matched-paired design. *Communications in Statistics-Theory and Methods* **37**, 2034-2051.

Sidhu, R., Hammon, B.G., Fuchs, R.L., Mutz, J.-N., Holden, L.R., George, B. and Olson, T. (2000). Glyphosate-Tolerant Corn: The Composition and Feeding Value of Grain from Glyphosate-Tolerant Corn Is Equivalent to That of Conventional Corn (Zean Mays L.). *Journal of Agricultural and Food Chemistry* **48**, 2305-2312.

Sileshi, G. (2006). Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological Research* **96**, 479-488.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D.(2007): *OpenBUGS User Manual*, Version 3.0.2, September 2007. http://www.math.helsinki.fi/openbugs/

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* **12(3)**, 1-16.

Sweeting, T.J. (2001). Coverage probability bias, objective Bayes and the likelihood principle. *Biometrika* **88**, 657-675.

Tamura, R.N. and Young S.S. (1986). The incorporation of historical control information in tests of proportions: Simulation study of Tarone's procedure. *Biometrics* **42**, 343-349.

Tarone, R.E. (1982). The use of historical control information in testing for a trend in proportions. *Biometrics* **38**, 215-220.

Venables, W.N., Ripley, B.D. (2002). *Modern Applied Statistics with S.* Springer, New York.

Venzon, D.J., and Moolgavkar, H. (1988): A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* **37**, 87-94.

Waller, R.A. and Duncan, D.B. (1969). A Bayes rule for the symmetric multiple comparison problem. *Journal of the American Statistical Association* **64**, 1484-1503.

Warton, D.I. (2005). Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics* **16**, 275-289.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association* **88**, 899-905.

Welch, B.L. and Peers, H.W. (1963). On formula for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society Series B* **25**, 318-329.

Wellek, S. (2003). Testing statistical hypotheses of equivalence. Chapman and Hall/CRC, Boca Raton.

Wellek, S. (2005): Statistical methods for the analysis of two-arm non-inferiority trials with binary outcomes. *Biometrical Journal* **47**, 48-61.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* **88**, 297-308.

Westfall, P.H., Johnson, W.O. and Utts, J.M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika* **84**, 419-427.

Westfall, P.H., Tobias, R.D., Rom, D., Wolfinger, R.D., Hochberg, Y. (1999). *Multiple Comparisons and Multiple Tests Using the SAS System*. SAS Institute Inc., Cary.

Westfall, P.H. and Young, S. (1993). *Resampling-Based Multiple Testing*. John Wiley and Sons, Inc., New York.

Williamson, P.P. (2006). Bayesian equivalence testing for binomial random variables. *Journal of Statistical Computation and Simulation* **77**, 739-755.

Wilson, E.B. (1927). Probable inference, the law of succession and statistical inference. *Journal of the American Statistical Association* **22**, 209-212.

Zhao, Y., Staudemayer, J., Coull, B.A., Wand, M.P. (2006). General design Bayesian generalized linear mixed models. *Statistical Science* **21**, 35-51.

Appendix A

Parametrization of distributions

In this Section, the distribution functions used in the text and models are are defined. Although this is standard knowledge for statisticians, these distributions are used with different parameterizations in Frequentist and Bayesian contexts and in the different software packages used in the model definitions for the simulation studies. Thence, to avoid confusion, the connections between the parameterizations are listed below.

A.1 The uniform distribution

The uniform distribution is throughout used with the parametrization $Y \sim unif(a, b)$:

$$f(Y) = \frac{1}{(b-a)}. \tag{A.1}$$

A.2 The normal (Gaussian) distribution

The pdf of a normal distributed random variable Y, $Y \sim N(\mu, \sigma^2)$ is:

$$f(Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \tag{A.2}$$

In this notation it is used throughout the text. In the Bayesian context, often the parametrization $Y \sim (\mu, \tau)$ is used instead, where τ is the precision parameter, $\tau = 1/\sigma^2$. This parametrization is also used in function `dnorm(mu, tau)` in BUGS models of Section .

A.3 The gamma distribution

The pdf of a random variable Y following a two-parameter gamma-distribution $Y \sim gamma\,(a, b)$ is defined:

$$f\,(Y) = \frac{y^{a-1}\exp\,(-y/b)}{b^a\Gamma\,(a)}, \tag{A.3}$$

with $y \geq 0$, $a > 0$, $b > 0$. Its first and second moment are $E\,(Y) = ab$ and $V\,(Y) = ab^2$. Further properties of the *gamma* distributions are described (Johnson et al., 1994). The above definition is equivalent to Johnson et al. (1994), Equation 17.23, with $a = \alpha$, and $b = \beta$. In this parametrization it will be used in the text. The `R`-function uses the parametrization `dgamma(x=x, shape=a, scale=b)`. In the Gibbs sampler implementations `WinBugs 1.4` and `OpenBUGS 3.0.2` in the function `dgamma(r, mu)` (Spiegelhalter et al., 2007) is defined with the parameters $r = a$ and $mu = 1/b$ according to the definitions above. Therefore, $E\,(Y) = r/mu$, $V\,(Y) = r/mu^2$.

A.4 The Poisson distribution

For random variable Y following Poisson distribution $Y \sim Pois\,(\mu)$, the probability of observing y is

$$P\,(Y = y) = \frac{\exp\,(-\mu)\,\mu^y}{y!}, \tag{A.4}$$

where the expectation of Y is $E\,(Y) = \mu$ and the variance of Y is $Var\,(Y) = \mu$. In this parametrization it is used in the text as well as in the code, where in `OpenBUGS` and `R`, `dgamma(lambda)` simply use `lambda=mu` in the definition above.

A.5 The negative binomial distribution

Many articles referring to overdispersion of count data assume negative binomial distribution. In this context, the negative binomial can be shown to arise from a mixture of Poisson distributions with means μ following a 2-parameter gamma distribution $gamma\,(a,b)$ with pdf:

$$f\,(\mu) = \frac{\mu^{a-1}\exp\,(-\mu/b)}{b^a\Gamma\,(a)}. \tag{A.5}$$

The probability to observe a certain count $Y = y$ arising from a mixture of Poissons with means distributed according to A.5 (Johnson et al., 1993) then is:

$$P\,(Y = y) = \left(\begin{array}{c} a+y-1 \\ a-1 \end{array}\right)\left(\frac{b}{b+1}\right)^y\left(\frac{1}{b+1}\right)^a. \tag{A.6}$$

In the literature the negative binomial distribution has been parameterized in many different ways. In the following, a notation in dependence of mean and dispersion parameter will be used, defining the probability to observe $Y = y$ in the notation $Y \sim NB\,(\mu,\tau)$ as:

$$P\,(Y = y) = \left(\begin{array}{c} \tau+y-1 \\ \tau-1 \end{array}\right)\left(\frac{\tau}{\mu+\tau}\right)^\tau\left(\frac{\mu}{\mu+\tau}\right)^y. \tag{A.7}$$

Here, $1/\tau$ is a dispersion parameter, and μ is the expectation of Y. In the negative binomial distribution, the variance is a quadratic function of the expectation and the dispersion parameter $1/\tau$: $Var\,(Y) = \mu + \frac{\mu^2}{\tau}$. As $\tau \to \infty$, the Poisson distribution results.

In the R-function `dbinom(x=x, size=n, prob=p)`:

$$P\,(X = x) = \frac{\Gamma\,(x+n)}{\Gamma\,(n)\,x!}p^n\,(1-p)^x\,, \tag{A.8}$$

with $a = (1-p)/p$, or $p = a/\,(1+a)$, $b = n$, $n > 0$, $E\,(X) = n\,(1-p)/p$, and $V\,(X) = n\,(1-p)/p^2$, when X follows a mixture of Poissons $Pois\,(\mu)$, with means $\mu \sim gamma\,(a,b)$.

The Gibbs sampler implementations `WinBugs 1.4` and `OpenBUGS 3.0.2` use an equivalent definition with in the function `dnegbin(p,r)`: $r = n$, $p = p$.

Additionally to the first, the `R` functions `dnbinom, rnbinom` used in the simulation study can use the alternative parameters: `dnbinom(x=x, size=n, mu=mu)`, using the following definitions and relations to the above parameters: $E(X) = \mu_{NB}$, $p = n/(n+\mu_{NB})$, $V(X) = \mu_{NB} + \frac{1}{n}\mu_{NB}^2$. The R package `gamlss` uses the same definition with $\sigma = 1/n = 1/\tau$.

Note: Data following the negative binomial distribution may also arise from counting the number of Bernoulli trials n that have to be performed until a certain prespecified number of successes $y = \sum Y$ is observed. Here, Y is random variable with values (0,1), being i.i.d. Bernoulli distributed $Y \sim Bern(\pi)$, with π being the probability to observe a success.

A.6 The multivariate normal distribution

An M-dimensional random variable $\boldsymbol{Y}^t = (Y_1, ..., Y_M)$ is denoted as M-variate normal distribution, $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, when is has a pdf

$$f(y_1, ..., y_M) = \frac{1}{\sqrt{(2\pi)^M |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right), \tag{A.9}$$

(Kotz et al., 2000), where $\boldsymbol{\mu}^t = (\mu_1, ..., \mu_M)$ is the vector of expected values and $\boldsymbol{\Sigma}$ is the $(M \times M)$ variance-covariance matrix, $\boldsymbol{y}^t$ denotes the transposed vector, Σ^{-1} denotes the inversion and $|\boldsymbol{\Sigma}|$ is the determinant of the variance-covariance matrix.

Appendix B

Basic R functions implementing the discussed methodology

In the following, basic R functions are presented, which implement basic, simple methodology. They allow to calculate simultaneous confidence intervals for multiple comparisons among elements primary parameter vector. They are sufficient to perform calculations leading to results as given for the examples in Sections 4.6.2 and 8.1. The functions are implemented in the R package `BSagri` which is available as a beta version from the author upon request.

B.1 SCI based on a joint empirical posterior distribution

The R code below implements the algorithm of Besag et al. (1995) outlined in Section 4.4 and is based on an initial version by Dilba (2006). The argument `x` must be an $(M \times K)$ matrix of K samples of the M dimensional parameter vector. The argument `conf.level` must be a single numeric value specifying the intended simultaneous confidence level, `alternative` must be a single character string, specifying whether two-sided intervals, upper or lower limits are to be contructed. The argument `...` is currently not used.

```
SCSnp.default <- function(x, conf.level=0.95, alternative="two.sided", ...){
alternative <- match.arg(alternative, choices=c("two.sided","less","greater"))
DataMatrix <- x
N <- nrow(DataMatrix)
k <- round(conf.level*N,0)
RankDat <- apply(DataMatrix,2,rank)

switch(alternative,
"two.sided"={
W1 <- apply(RankDat,1,max)
W2 <- N + 1 - apply(RankDat,1,min)
Wmat <- cbind(W1,W2)
w <- apply(Wmat,1,max)
tstar <- round(sort(w)[k],0)
SCI <- function(x){sortx <- sort(x)
 cbind(sortx[N+1-tstar],sortx[tstar])}
SCS <- t(apply(DataMatrix,2,SCI))},

"less"={
W1 <- apply(RankDat,1,max)
tstar <- round(sort(W1)[k],0)
SCI <- function(x){sortx <- sort(x)
 cbind(-Inf, sortx[tstar])}
SCS<-t(apply(DataMatrix,2,SCI))},

"greater"={
W2 <- N + 1 - apply(RankDat,1,min)
tstar <- round(sort(W2)[k],0)
SCI <- function(x){sortx <- sort(x)
 cbind(sortx[N+1-tstar], Inf)}
SCS<-t(apply(DataMatrix,2,SCI))
})

estimate<-apply(DataMatrix,2, median)
colnames(SCS)<-c("lower","upper")
out<-list(
conf.int=SCS, estimate=estimate, x=x, k=k, N=N,
```

```
conf.level=conf.level, alternative=alternative)
class(out)<-"SCSnp"
return(out)}
```

B.2 Joint empirical posterior of multiple contrasts

The following R code implements the functions to compute contrasts of interest based on differences and ratios of a primary parameter vector. The argument `x` must be a $(K \times I)$ matrix of K samples of the primary parameter vector, and the argument `cmat` must be an $(M \times I)$ matrix. This function implements Equation (4.21).

```
CCDiff.default <-function(x, cmat){
require(multcomp)
if(!is.matrix(x) & !is.data.frame(x))
 {stop("Argument 'x'must be a matrix or data.frame!")}
ngroup<-ncol(x)
Nsim<-nrow(x)
chains<-x
if(!is.matrix(cmat))
 {stop("'cmat' must be a matrix, specifying the contrast coefficients")}
if(ngroup!=ncol(cmat))
 {stop("ncol(cmat) must be the same as the number of means in muvec")}
cs<-apply(cmat,1,sum)
if(any(cs!=0))
 {warning("Rows of cmat do not sum up to zero.
  Are the contrasts appropriately defined?")}

nchains<-apply(X=chains, MARGIN=1, FUN=function(x){cmat %*% x})
if(nrow(cmat)==1)
 {nchains<-matrix(nchains, nrow=1)}
rownames(nchains)<-rownames(cmat)
out<-list(
chains=t(nchains), x=x, cmat=cmat)
class(out)<-"CCDiff"
```

```
return(out)
}
```

The following function implements Equation (4.22). The argument `x` must be a $(K \times I)$ matrix of K samples of the primary parameter vector, and the argument `cmat` must be a list with elements `numC` and `denC` both being $(M \times I)$ matrices.

```
CCRatio.default <- function(x, cmat){
require(mratios)
ngroup<-ncol(x)
chains<-x
if(!is.list(cmat))
 {stop("cmat must be a list")}
if(is.null(cmat$numC)|is.null(cmat$denC))
 {stop("cmat must be a list with elements $numC and $denC,
  specifying the numerator and denominator contrast coefficients")}
if(!is.matrix(cmat$numC)|!is.matrix(cmat$denC))
 {stop("elements $numC and $denC of 'cmat' must be matrices,
  specifying the numerator and denominator contrast coefficients")}
if(ngroup!=ncol(cmat$numC))
 {stop("ncol(cmat$numC) must be the same as the number of means in muvec")}
if(ngroup!=ncol(cmat$denC))
 {stop("ncol(cmat$denC) must be the same as the number of means in muvec")}

nchains<-apply(X=chains, MARGIN=1, FUN=function(x){(cmat$numC%*%x) / (cmat$denC%*%x)})
if(nrow(cmat$numC)==1)
 {nchains<-matrix(nchains, nrow=1)}
rownames(nchains)<-rownames(cmat$numC)
out<-list(
chains=t(nchains), x=x, cmat=cmat)
class(out)<-"CCRatio"
return(out)
}
```

Acknowledgements

The project underlying this work was funded by Bundesministerium für Bildung und Forschung, grant number 0313269. The responsibility for the content lies with the author.

I thank Prof. Dr. Ludwig A. Hothorn for providing the project, statistical problems and seminal ideas for the solutions, for his constant assistance and education, Dr. Gemechis Dilba Djira for providing seminal ideas and approaches, as well as helpful comments on a related internal report, Prof. Hans-Peter Piepho for his detailed comments and corrections on an earlier version which helped to make this work more precise, readable and understandable, Mario Hasler for his patience in teaching me the basics of mathematical notation and LaTeX, Mario Hasler and Daniel Gerhard for all the lively discussions on multiple comparisons, joint distributions and generalized linear models, Stephan Rauschen and Dr. Sabine Prescher for providing the practical problems and data sets. Further, I thank Hanne Visser and Clemens Buczilowski for helping me to administrate my life as an employee and computer user. Last, but not least, I thank my parents, Ruth, my brother and sisters and the ESG Hannover for helping me to keep balance.