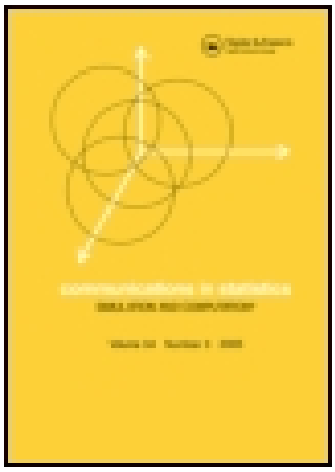

Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Simultaneous Small Sample Inference For Linear Combinations Of Generalized Linear Model Parameters

Daniel Gerhard[a]

[a] Institute of Biostatistics, Leibniz Universität Hannover, Hannover, Germany,

Accepted author version posted online: 19 Jun 2014.

To cite this article: Daniel Gerhard (2014): Simultaneous Small Sample Inference For Linear Combinations Of Generalized Linear Model Parameters, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2014.895836

To link to this article: http://dx.doi.org/10.1080/03610918.2014.895836

SIMULTANEOUS SMALL SAMPLE INFERENCE FOR LINEAR COMBINATIONS OF GENERALIZED LINEAR MODEL PARAMETERS

Daniel Gerhard

Institute of Biostatistics

Leibniz Universitt Hannover

Hannover, Germany

gerhard@biostat.uni-hannover.de

Key Words: Simultaneous confidence intervals; Generalized linear models; Modified likelihood root; General linear hypotheses; Multiple contrast test.

ABSTRACT

A method is proposed to construct simultaneous confidence intervals for multiple linear combinations of generalized linear model parameters, that uses a multivariate normal- or t-distribution together with the signed likelihood root statistic. In an application to a case study simultaneous confidence bands for logistic regression are calculated. A simulation study based on the example evaluation suggests superior performance compared to the common Wald-type approaches. The proposed methods are readily implemented in the R extension package mcprofile.

1. INTRODUCTION

Generalized linear models (McCullagh and Nelder, 1989) are a well established family of models with a wide range of application. With a sample of observations under assumption of a distribution from the exponential family and a set of explanatory covariates, parameters can be estimated and predictions for a new set of input variables can be obtained. For example, logistic regression or models for count data are prominent areas of application, assuming either a Binomial or a Poisson distributed response, or a generalization of these distributions with additional parameters in the variance function. When doing inference for parameters of a generalized linear model, one

can rely on large sample approximations if sufficiently large data is available. To provide accurate inference at small sample sizes, profile likelihood methods and higher order asymptotics (Brazzale and Davison, 2008) are a prominent way to construct confidence intervals for a single parameter in the model.

In this article the focus is set on inference based on a set of profile statistics, controlling the family-wise error rate (FWER), that is, the probability of falsely rejecting at least one true null hypothesis, at a specified level. Instead of providing adjusted p-values and simultaneous confidence intervals directly for the model parameters, inference for derived parameters is considered, specifying linear combinations of parameters by providing a matrix of contrast coefficients.

2. GENERALIZED LINEAR MODELS

First, a brief overview of parameter estimation in generalized linear models is given. For more detailed information the reader is referred e.g. to the book of McCullagh and Nelder (1989). A vector of $i = 1, \ldots, n$ observations $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ is assumed to be a realization of a random variable $\boldsymbol{Y}$, where each component of $\boldsymbol{Y}$ is assumed to have a distribution in the exponential family. The systematic component of a generalized linear model is defined as

$$g() = \boldsymbol{\eta} = \boldsymbol{X\beta}$$

with a link function $g(\cdot)$, a p-dimensional vector of parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, and a (np) matrix $\boldsymbol{X} = (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T)^T$ with n row vectors of design covariates for each observation. Given the vector of predictor variables, the log-likelihood can be written as the sum of the logarithmic density function evaluated at each of the n observations $l(; \boldsymbol{y}) = \sum_{i=1}^{n} \log f_i(y_i; _i)$. Instead of the likelihood function, the scaled deviance

$$D(\boldsymbol{y};) = 2l(\boldsymbol{y}; \boldsymbol{y}) - 2l(; \boldsymbol{y})$$

can be used as a goodness-of-fit criterion. To estimate a coefficient vector $\hat{\boldsymbol{\beta}}$ an iteratively reweighted least squares (IRWLS) algorithm can be applied (McCullagh and Nelder, 1989) for find-

ing the minimum of the deviance function.

In many applications the experimental questions are specified through $k = 1, \ldots, q$ linear combinations of the model parameters, $\boldsymbol{\vartheta} = \boldsymbol{A}\boldsymbol{\beta}$, which are defined by a (qp) contrast matrix $\boldsymbol{A} = (\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_q^T)^T$, where each of the q row-vectors $\boldsymbol{a}_k$ contains predefined constants to define a single contrast parameter ϑ_k. When constructing simultaneous confidence intervals or hypotheses tests, the multiple comparison problem of testing all hypotheses at a nominal level of α, the overall type I error rate, has to be considered. A unifying simultaneous inference framework for these linear combinations of parameters in general parametric models is presented in Hothorn et al. (2008). They consider the general linear hypothesis (Searle, 1971, p.110):

$$H_0: \quad \boldsymbol{\vartheta} = \boldsymbol{m}$$

where $\boldsymbol{m} = (m_1, \ldots, m_q)$ is a vector of specified constants defining the test margins. This global hypothesis is partitioned into the q different sub-hypotheses, testing each ϑ_k separately, but maintaining the global type-I-error rate. The key factor of this single-step inference is the assumption of a multivariate normal-distribution of the standardized estimator $\hat{\boldsymbol{\vartheta}}$ with a correlation structure, which is directly obtained from the (pp) observed information matrix at the parameter estimates $j(\boldsymbol{\beta}) = -\frac{\partial^2 l(;\boldsymbol{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$.

3. TEST STATISTICS

A single element ϑ_k of the vector $\boldsymbol{\vartheta}$, corresponding to a single row $\boldsymbol{a}_k = (a_1, \ldots, a_q)$ of the (qp) contrast matrix $\boldsymbol{A}$ is used to introduce different test statistics to test an elementary null-hypothesis.

In Hothorn et al. (2008) the Wald-type statistic

$$w(\vartheta_k) = \left(\hat{\vartheta}_k - \vartheta_k\right) j^{\frac{1}{2}}(\hat{\vartheta}_k)$$

is used, where $j^{-1}(\hat{\vartheta}_k) = \boldsymbol{a}_k j^{-1}(\hat{\boldsymbol{\beta}}) \boldsymbol{a}_k^T$ is the inverse of the observed information of the contrast parameter. As the observed information is fixed at the maximum likelihood estimates for any ϑ_k,

the statistic $w(\vartheta_k)$ is a linear function, and therefore approximates the deviance by a quadratic function $w^2(\vartheta_k)$, which holds exactly for a Gaussian linear model. For a non-Gaussian response the quadratic approximation might be inadequate at small sample sizes, especially if the parameter space is bounded. Hence, the performance of the Wald-type statistic highly depends on the choice of the link function, as lower and upper confidence interval limits based on this statistic will have equal distance to the maximum likelihood estimate on the scale of the linear predictor.

To improve the asymptotic properties of the test, the signed root deviance statistic (Chen and Jennrich, 1996)

$$q(\vartheta_k) = sign\left(\hat{\vartheta}_k - \vartheta_k\right)\sqrt{\frac{D(\boldsymbol{y};\hat{}) - D(\boldsymbol{y};\tilde{})}{\phi}}$$

can be applied as an alternative to the Wald-type statistic. $\hat{}$ is the linear predictor at the maximum likelihood estimates of the parameters $\hat{\boldsymbol{\beta}}$. $\tilde{}$ denotes the linear predictor at the restricted parameter estimates $\tilde{\boldsymbol{\beta}}$, obtained under the linear constraint $\boldsymbol{a}_k\boldsymbol{\beta} = \vartheta_k$.

ϕ is a dispersion parameter, which accounts for extra variation in the data. This parameter is fixed at 1 e.g. for a Binomial or Poisson model, but can also be estimated from the data, like the residual error in a Gaussian linear model. As the deviance function is an essential part of the test statistic in comparison of just using the quadratic approximation, improving this approximation comes at the cost, that the model has to be refitted several times to obtain the deviance values in the neighborhood of the maximum likelihood estimate. These model updates are additionally complicated, as interest lies in the derived parameters $\boldsymbol{\vartheta}$ instead of the parameter vector $\boldsymbol{\beta}$.

To obtain$\tilde{}$, the weighted least squares step in the IRWLS algorithm can be modified by using a weighted regression, which allows to apply linear equality constraints on the regression parameters at each iteration. A quadratic programming algorithm, e.g. the dual method by Goldfarb and Idnani (1983), can be used to obtain the restricted parameter estimates. An application of a related algorithm for fitting shape constraint generalized linear models is presented in Meyer (2012).

For higher order density approximations, a modified likelihood root is given in Barndorff-

Nielsen (1983), which can be extended to general linear hypotheses by

$$r(\vartheta_k) = q(\vartheta_k) + \frac{1}{q(\vartheta_k)} \log\left(\frac{w(\vartheta_k)\rho(\vartheta_k, \hat{\vartheta}_k)}{q(\vartheta_k)}\right),$$

with $\rho(\vartheta_k, \hat{\vartheta}_k) = \sqrt{\frac{|j_{\lambda\lambda}(\hat{\boldsymbol{\beta}})|}{|j_{\lambda\lambda}(\tilde{\boldsymbol{\beta}})|}}$. $|j_{\lambda\lambda}(\cdot)|$ denotes the determinant of a subset of the observed information matrix to summarize the information about the nuisance parameters that are not subject to the linear constraint. In terms of inference for a contrast parameter, the nuisance parameters correspond to contrast coefficients of zero. Additional to the likelihood root, the observed information is evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ and the restricted estimates $\tilde{\boldsymbol{\beta}}$ under the linear constraint $\boldsymbol{a}_k\boldsymbol{\beta} = \vartheta_k$. The observed information matrix for the equality-constraint parameters can be obtained, according to Liew (1976), by

$$j(\tilde{\boldsymbol{\beta}}) = \left(\widehat{\boldsymbol{M}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{M}}^T\right)^{-1}, \quad \widehat{\boldsymbol{M}} = \boldsymbol{I} - \widehat{\boldsymbol{\Sigma}}\boldsymbol{a}_k^T\left(\boldsymbol{a}_k\widehat{\boldsymbol{\Sigma}}\boldsymbol{a}_k^T\right)^{-1}\boldsymbol{a}_k, \quad \widehat{\boldsymbol{\Sigma}} = j^{-1}(\hat{\boldsymbol{\beta}}).$$

As the variance-covariance matrix of the restricted parameters might be singular due to the equality constraints, an eigenvalue or singular value decomposition might be used to compute the determinant of interest, based on eigenvalues larger than zero. A further difficulty arises, as the statistic $r(\vartheta_k)$ is not defined at the maximum likelihood estimates.

4. SIMULTANEOUS INFERENCE

We now consider to test each of the q null hypotheses individually under control of the family-wise error rate, either using the statistics $w(\vartheta_k)$, $q(\vartheta_k)$, or $r(\vartheta_k)$. To maintain the FWER, the global null hypothesis is rejected, if at least one elementary hypothesis is rejected; thus, focus is set on the maximum of test statistics. The distribution of this maximum evaluated at the specific test margins m_k can be specified for a two-sided testing procedure as

$$P(\max|w(m_k)| \leq t) \cong \int_{-t}^{t} \cdots \int_{-t}^{t} \varphi\left(x_1, \ldots, x_q; \boldsymbol{R}, \nu\right) dx_1 \ldots dx_q = g_\nu(\boldsymbol{R}, t)$$

for any $t \in \mathbb{R}$. φ is either the multivariate normal- or t-distribution function, given a residual error degree of freedom ν, which is assumed when the dispersion parameter ϕ is estimated from the

data. The correlation structure $\boldsymbol{R}$ is obtained by standardizing the variance-covariance matrix of the contrast parameters,

$$\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{D}}^{-1/2}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{D}}^{-1/2}, \quad \text{where} \quad \widehat{\boldsymbol{D}} = diag\left(\widehat{\boldsymbol{\Psi}}\right)\boldsymbol{I}_q, \quad \text{and} \quad \widehat{\boldsymbol{\Psi}} = \boldsymbol{A}\,j^{-1}(\hat{\boldsymbol{\beta}})\boldsymbol{A}^T$$

using the estimates from the data and treat them as if it were the true correlation matrix. Efficient approaches to approximate these multiple integrals are discussed in Bretz et al. (2001).

Under assumption of a linear model with normal distributed residuals, the signed root deviance statistics equal the Wald-type statistic (Bates and Watts, 1988, p. 205), leading to exact statistical inference as the correlation structure of the multivariate *t*-distribution is defined only by the known contrasts and design covariates. In a more general setting, a second-order approximation to the deviance is used (Lindsey, 1996, p. 209), calculating the correlation structure, using the observed information at the maximum likelihood estimates similar to the Wald-type statistic. The marginal distributions of the q test statistics follow asymptotically a standard normal or *t*-distribution (Brazzale and Davison, 2008), hence the approximated correlation structure will only influence the degree to which the tests are adjusted for multiplicity.

Adjusted p-values controlling the FWER are calculated as

$$p_k = 1 - g_\nu(\widehat{\boldsymbol{R}}, |t_k|)$$

where t_k is the observed statistic, either using $w(m_k)$, $q(m_k)$, or $r(m_k)$.

Instead of using adjusted p-values, confidence intervals are defined by inverting the hypothesis test as

$$I = \{\vartheta_k : -c_{1-\alpha} \leq w(\vartheta_k) \leq c_{1-\alpha}\}$$

with a critical value $c_{1-\alpha}$. Analogously, $w(\vartheta_k)$ can be substituted by the different profile statistics $q(\vartheta_k)$ and $r(\vartheta_k)$. As the confidence limits are found separately for each of the q linear combinations of parameters, the combined set of confidence limits describes a rectangular confidence set in a q-

dimensional space.

The critical value $c_{1-\alpha}$ should be chosen in a way that the FWER is controlled, considering the correlation between the derived parameters. Similar to the testing procedure, the control of the global error rate is maintained by focusing on the maximum of the q elementary hypotheses, where the statistics, analogous to Hothorn et al. (2008), are assumed to follow a q-variate normal or t-distribution. In order to assign the same weight or error level to each of the q hypotheses an equicoordinate quantile $c_{1-\alpha}$ is calculated from this q-dimensional distribution (Bretz et al., 2001).

For an effective search for the confidence limits, the strategy of Venables and Ripley (2002, p. 221) is adopted, establishing a grid of values for each of the ϑ_k around the maximum likelihood estimates, and interpolating the resulting $q(\vartheta_k)$ or $r(\vartheta_k)$ by a cubic spline function. The confidence limits are found by evaluating the inverse of this interpolating function at $-c_{1-\alpha}$ and $c_{1-\alpha}$. As an alternative, a simple bisection method can be utilized to search for each confidence limit directly, but missing the opportunity of gaining additional insights by a graphical representation of the profiled parameter.

5. APPLICATION TO A CASE STUDY

In a dose-response experiment the lethal effect of an insecticide is tested. The data example is artificially generated, based on an excerpt of a real experiment, featuring very small sample sizes. The generated data is shown in Table I.

5.1 ESTIMATING THE LETHAL DOSE

The objective in this experiment is the detection of a lethal dose of the insecticide $LD(p)$, affecting a specific fraction p [%] of the tested subjects. The dose-response curve is modeled by a logistic regression model with

$$y_i \sim Binomial(n_i, {}_i), \qquad \eta_i = g({}_i) = \log\left(\frac{{}_i}{1-{}_i}\right), \qquad \eta_i = \beta_1 + \log(x_i)\beta_2.$$

The $LD(p)$ can be obtained by inverse regression, estimating the dose level, which corresponds

to the linear predictor at the cutoff fraction p. A confidence interval for the $LD(p)$ is obtained by searching for the dose corresponding to the confidence limits of the linear predictor at the same cutoff value. Therefore, a first step will be the calculation of simultaneous confidence bands for the logistic regression curve, which can be obtained by specifying a suitable contrast matrix.

The transformed parameter vector $\boldsymbol{\vartheta} = \boldsymbol{A}\boldsymbol{\beta}$ will represent predictions for new, unobserved dose levels by choosing a contrast matrix $\boldsymbol{A}$ which resembles a design matrix for a new dataset with column vectors with design coefficients for the intercept a_{k1} and for pre-specified dose levels a_{k2}. The $LD(p)$ and corresponding confidence limits are found by searching for the dose levels at which $g^{-1}(\vartheta) = p$ and the projection of corresponding confidence limits for neighboring ϑ at level $g(p)$. These projections of cutoff intersections are illustrated in Figure 1 for the lower confidence limit of the $LD(25)$.

With a large number of rows already in the contrast matrix $\boldsymbol{A}$, corresponding to a dense grid of dose levels a_{k2}, only a marginal change in the global type-I-error rate can be expected with a further increase of the density of the grid, as the test statistics based on two neighboring dose levels can be assumed to be highly correlated. Hence, the error rate of falsely rejecting at least one null-hypotheses corresponding to any $LD(p)$ within the range of a_{k2} can be controlled by specifying an adequate grid of dose levels covering the dose range of interest. The effect of controlling the FWER with the proposed plugin-method in comparison to the control of the comparison-wise error rate and a common Bonferroni adjustment is presented in Figure 2. The proposed method results in confidence limits with only a small distance to the unadjusted confidence limits, whereas the Bonferroni adjustment results in much wider intervals.

The estimated $LD(p)$ at $p \in \{25, 50, 75\}$ are found at $\{1.81, 2.89, 4.60\}$. The lower simultaneous confidence limits for these parameters can be found in Table II. Especially at small p, the profiling methods show a smaller distance of the lower limit to the point estimate. In this case, the higher order approximations can be seen as a compromise between Wald-type and the first order profile confidence intervals.

5.2 MULTIPLE COMPARISONS TO A CONTROL

To illustrate the advantages and the limits of the profile methods, multiple comparisons of the success rates at each dose level to the control dose are performed. A similar model as in Section 5.1 is assumed, with the linear predictor $\eta_i = \beta_j$, where $j = 1, \ldots, n$, estimating the success rates individually for each dose level x_i. In this model the distances between each dose level is not considered, as the design matrix is just the identity matrix.

Simultaneous confidence intervals and tests are calculated based on the contrasts

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

with comparisons of rates at each dose level to the control dose. Simultaneous confidence intervals for ϑ_k and multiple tests of the null hypotheses $H_0^{(k)}: \ \vartheta_k = 0$ are provided.

The problem for the particular data at hand is the success rate of 0 out of 10 at the second dose level and 10 out of 10 at the last dose. Adding the number of non-zero contrast coefficients divided by the number of model parameters, that is $2/6$, as pseudo-events to the successes and failures in each dose group, according to Price and Bonnett (2004), allows to make inference about the ϑ_k in spite of observing unadjusted rates at the border of the parameter space.

The estimated lower and upper simultaneous confidence limits and the adjusted p-values are shown in Table III. When comparing rates at higher dose levels to the control, the profile method obtains smaller p-values and likewise compatible lower confidence limits with a larger distance to zero.

6. SIMULATION STUDY

To evaluate the performance of the methods, the simultaneous coverage probability of the in-

tervals is examined by simulation. The simulation settings are related to the data example, performing a logistic regression with support at $x_i = (0, 1, 2, 3, 4)$, parameterized by $\beta_1 = -2.197$ and $\beta_2 = 1.099$ with $g(_i) = \eta_i = \beta_1 + x_i\beta_2$, resulting in a vector $_i = (0.1, 0.25, 0.5, 0.75, 0.9)$. Data is generated from a Binomial distribution, $y_i \sim Binomial(n_i, _i)$, for different numbers of $n_i \in \{3, 4, \ldots, 25\}$. At each sample size setting n_i, 100,000 simulation runs are performed. At extremely small sample sizes, the constrained estimation algorithm might not converge; when no confidence limit can be obtained at a simulation run, this limit is fixed at the border of the parameter space.

In a first part of the simulation, the coverage probability of simultaneous $(1 - \alpha) = 0.95$ confidence intervals for $g^{-1}(\vartheta_k)$ with $\boldsymbol{\vartheta} = \boldsymbol{A\beta}$ is examined, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

The simulation results are shown in Figure 3. The Wald-type intervals are showing a conservative behavior, especially at small sample sizes, whereas the signed root deviance profile results in anti-conservative intervals, but reaching the nominal level a bit faster than the Wald approach with increasing sample sizes. The best performance is given by the higher order approximation; only at $n_i \leq 5$ some numerical problems occur, which are influencing the simulation results.

Instead of confidence intervals for $_i$, the confidence intervals for $LD(p)$ are of interest in the data example. In order to evaluate the performance of these intervals a second simulation study is conducted. Rather than calculating confidence intervals at support coordinates of $a_{k2} = (1, 2, 3)$, a whole range of 50 equally spaced coordinates from -10 to 15 are chosen ($a_{k2} = (-10, -9.49, \ldots, 15)$). The coverage probability of simultaneous confidence intervals is simulated for 50 $LD(p)$ parameters in a range between 1 and 3 based on 10,000 runs at each n_i setting. The simulation results are presented in Figure 4. Due to numerical instabilities at the border of the pa-

rameter space, the simulated coverage probabilities are far below the nominal level for all methods at very small sample sizes. But at around $n_i \geq 20$ the coverage probabilities converge at a level slightly higher than the nominal level. As a single error rate is controlled for nearly all $LD(p)$ parameters within a certain range, all approaches show a slightly conservative behavior; hence the in other respects anti-conservative performance of the profile approach is beneficial in this certain situation.

To investigate the performance of comparisons to a control, corresponding to the application in Section 5.2, the $_i$ related to 6 different dose groups are generated from a uniform distribution $_i \sim U(0,1)$; hence, a wide range of different response profiles are summarized within 100,000 simulation runs. In a second simulation the $_i$ are sampled from a Beta distribution $_i \sim Beta(5,5)$, omitting extreme $_i$ at the border of the parameter space. Analogously to the previous simulations, the response vector is generated by $y_i \sim Binomial(n_i, _i)$, for different numbers of $n_i \in \{5, 10, \ldots, 100\}$. The response vector is adjusted for small sample sizes by adding $2/6$ pseudo-events to the successes and failures in each dose group according to Price and Bonnett (2004).

When generating the group level proportions from a $U(0,1)$ distribution, the likelihood root methods show similar characteristics as the Wald-type statistic (Figure 5). Due to computational instabilities, when estimating the profile statistics at small $_i$, no reliable confidence limits can be obtained; hence, the limits are set to $[-\infty, \infty]$, resulting in conservative coverage properties. This problem at extreme $_i$ is pointed out, when omitting the problematic parameter region by sampling from the $Beta(5,5)$ distribution. The results in Figure 6 show, that both likelihood root methods are reaching the nominal coverage level much faster with increasing sample sizes compared to the Wald-type statistic.

7. DISCUSSION

At small sample sizes, the use of likelihood profiles can improve the properties of simultaneous confidence intervals compared to the Wald-type approach, proposed by Hothorn et al. (2008). Es-

pecially, at the border of the parameter space, more accurate and useful inference can be obtained, at which the Wald-type intervals just remain completely uninformative. As the Wald approach relies on a quadratic approximation to the log-likelihood function, an adequate choice of the link function will certainly improve the approximation by a quadratic function. As a further advantage of the profile statistics, the performance of the confidence intervals is not as dependent on an adequate choice of a link function compared to the Wald approach, as the log-likelihood function is a major part of the profile statistic itself. Only the calculation of the critical value $c_{1-\alpha}$ is based on the estimated covariance matrix of the model parameters, which is dependent on the chosen link function.

A certain disadvantage is the increased computational effort, as the construction of a profile requires additional constrained parameter estimation steps. Together with the search for the equicoordinate quantile $c_{1-\alpha}$ this might increase the computation time, dependent on the complexity of the model and the number of contrast parameters (only dimensions $\leq 1,000$ available). A software implementation is available as an R package mcprofile (`http://cran.r-project.org/web/packages/mcprofil`) using the existing generalized linear modeling functions in R.

BIBLIOGRAPHY

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

Bates, D. M. and Watts D. G. (1988). *Nonlinear Regression Analysis and Its Applications.* Wiley Series in Probability and Statistics; John Wiley & Sons, Inc., Hoboken, New Jersey.

Brazzale, A. R. and Davison, A. C. (2008). Accurate parametric inference for small samples. *Statistical Science*, **23**, 465–484.

Bretz, F., Genz, A., and Hothorn, L. A. (2001). On the numerical availability of multiple comparison procedures. *Biometrical Journal*, **43**, 645–656.

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

Goldfarb, D. and Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, **27**, 1–33.

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

Liew, C. K. (1976). Inequality constrained least-squares estimation. *Journal of the American Statistical Association*, **71**, 746–751.

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

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC.

Meyer, M. C. (2012). A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics - Simulation and Computation*, **42**, 1126–1139.

Price, R. M. and Bonett, D. G. (2004). An improved confidence interval for a linear function of binomial proportions. *Computational Statistics & Data Analysis*, **45**, 449–456.

Searle, S. R. (1971). *Linear Models*. Bd. 1 in Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley & Sons, New York.

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

Table I: Dose-response data, investigating the mortality of insects with increasing dose levels of an insecticide.

Dose (x_i)	No. of dead insects (y_i)	Total (n_i)
0	1	10
0.625	0	9
1.25	1	10
2.5	2	10
5	9	10
10	10	10

Table II: $LD(p)$ simultaneous lower confidence limits, controlling the FWER at a type-I-error level of $(1 - \alpha) = 0.95$.

	p [%]		
Method	25	50	75
likelihood root	0.97	1.74	3.27
Wald-type	0.71	1.73	3.01
modified lik. root	0.94	1.93	3.30

Table III: Simultaneous confidence intervals and adjusted p-values for comparing rates for each dose to the control on the logit link. At missing entries no confidence limit could be calculated.

		Wald-type			likelihood root			modified likelihood root		
comparison	estimate	lower	upper	p-value	lower	upper	p-value	lower	upper	p-value
0.625 - 0	-1.38	-6.43	3.66	0.95		3.10	0.92		3.19	0.96
1.25 - 0	0.00	-3.32	3.32	1.00	-3.88	3.88	1.00	-3.76	3.76	1.00
2.5 - 0	0.67	-2.33	3.68	0.98	-2.35	4.42	0.97	-2.35	4.25	0.98
5 - 0	3.89	0.58	7.21	0.01	1.08	8.17	< 0.01	1.01	7.95	< 0.01
10 - 0	5.38	0.34	10.42	0.03	1.78		< 0.01	1.64		< 0.01

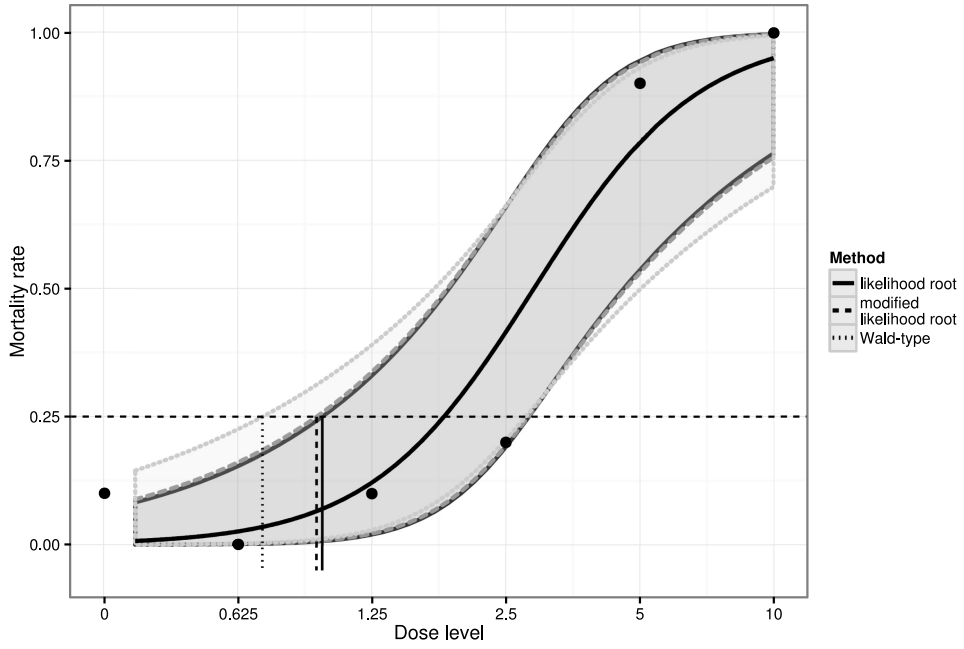

Figure 1: Pointwise simultaneous confidence bands for the logistic regression example and LD(25) estimates.

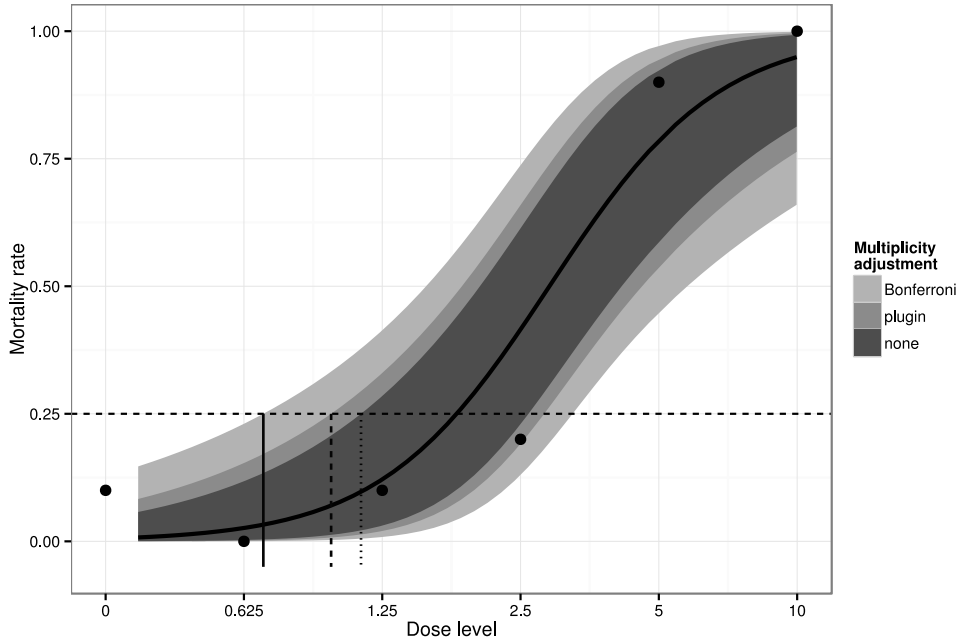

Figure 2: Comparing methods for multiplicity adjustment of pointwise simultaneous confidence bands for the logistic regression example.

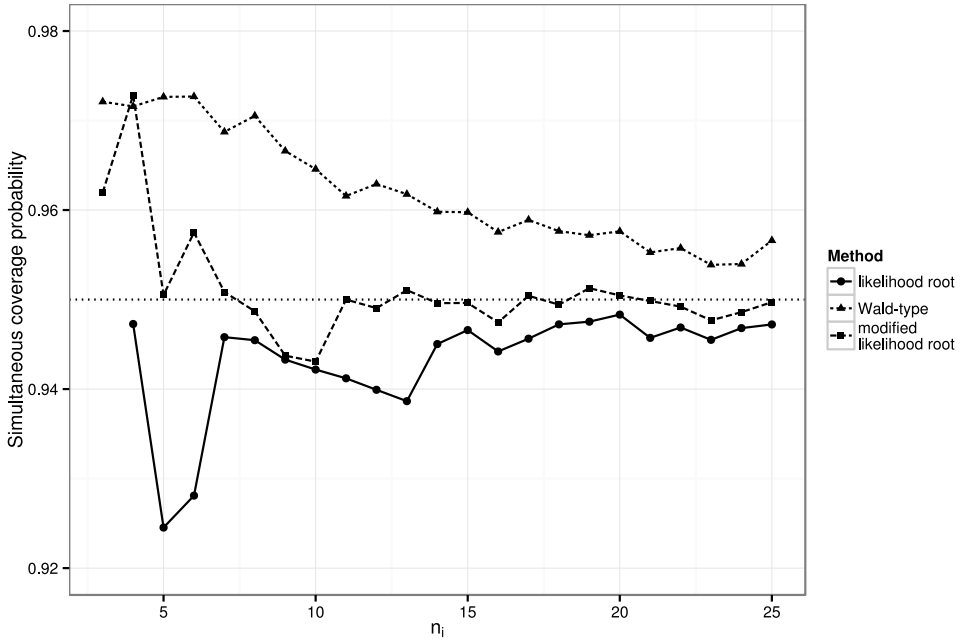

Figure 3: Simulated coverage probabilities of confidence intervals for specific $?_i$ with increasing sample sizes n_i.

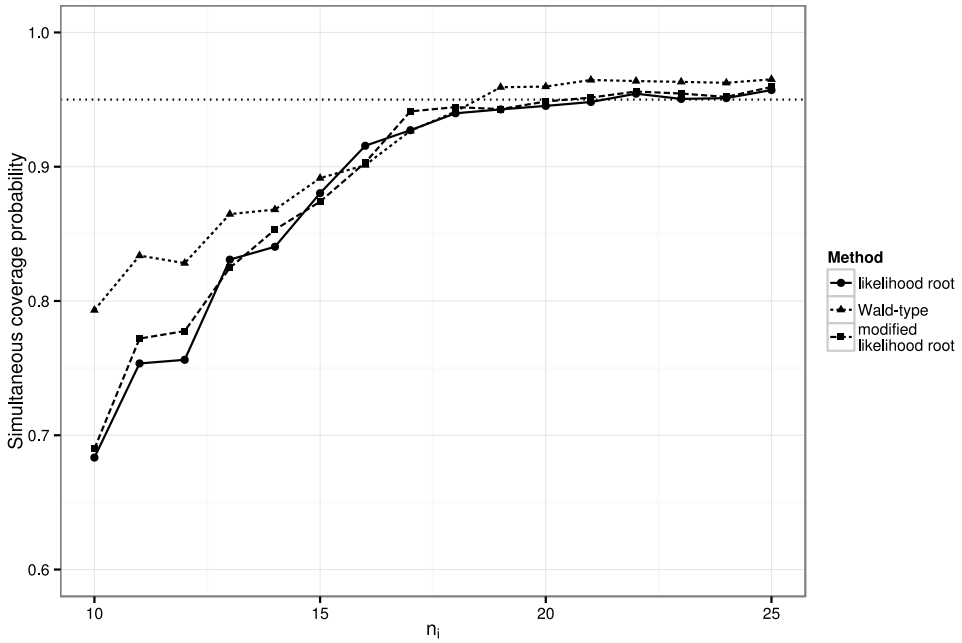

Figure 4: Simulated coverage probabilities of confidence intervals for specific LD(p) with increasing sample sizes n_i.

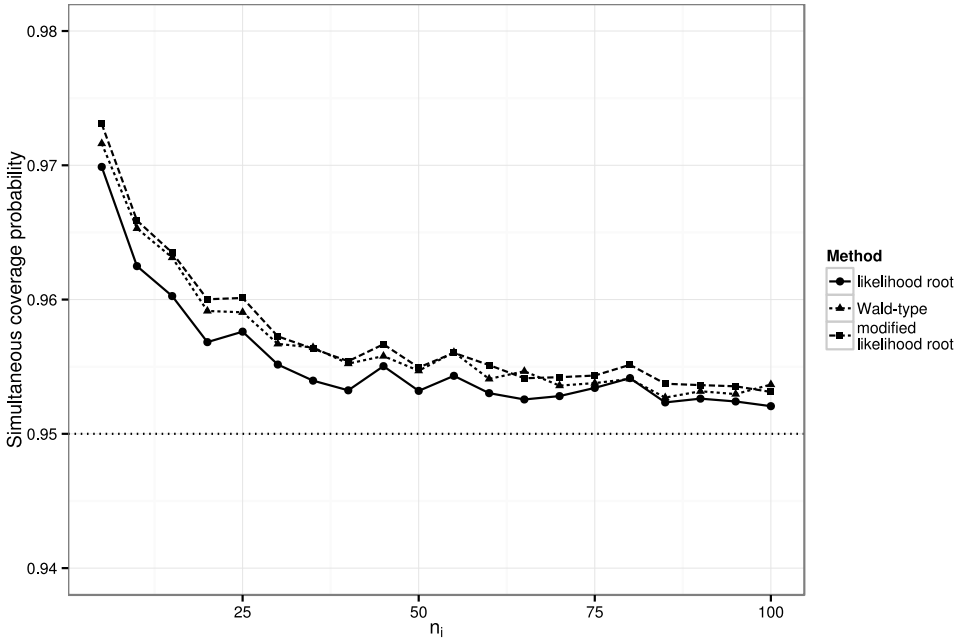

Figure 5: Simulated coverage probabilities of confidence intervals for comparisons to a control with increasing sample sizes n_i, generating rates from a U(0,1) distribution.

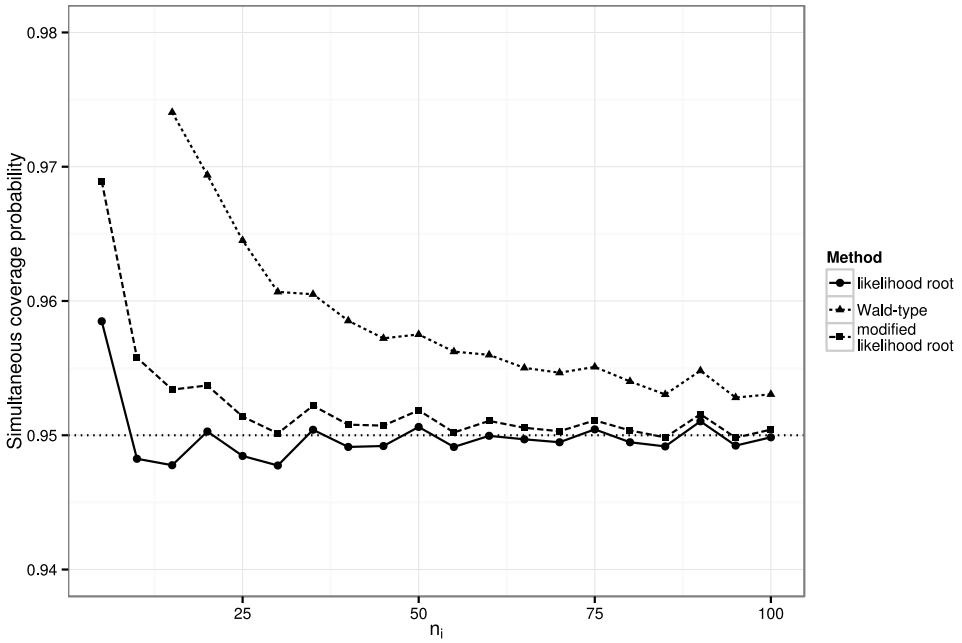

Figure 6: Simulated coverage probabilities of confidence intervals for comparisons to a control with increasing sample sizes n_i, generating rates from a Beta(5,5) distribution.