

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per le Tecnologie e le Scienze



Relazione Finale  
Likelihood-based Inference for  
Linear-by-Linear Association in  
Contingency Tables

Relatore Prof. Nicola Sartori  
Dipartimento di Scienze Statistiche

Laureando: Giorgia Bergamo  
Matricola N. 1100335

Anno Accademico 2016/2017



# Contents

<b>Introduction</b>	<b>5</b>
<b>1 Likelihood</b>	<b>7</b>
1.1 Likelihood Function . . . . .	7
1.2 Maximum Likelihood . . . . .	8
1.3 Regularity Conditions . . . . .	9
1.4 Score Function . . . . .	9
1.5 Information . . . . .	10
1.5.1 Observed Information . . . . .	10
1.5.2 Expected information . . . . .	10
1.6 Log-Likelihood Ratio Statistic . . . . .	11
1.7 Signed Root Log-likelihood Ratio Statistic . . . . .	11
1.8 Profile Likelihood . . . . .	12
<b>2 Loglinear Models for Contingency Tables</b>	<b>15</b>
2.1 Models for Count Data . . . . .	15
2.1.1 Poisson Model . . . . .	16
2.1.2 Loglinear Models for Contingency Tables . . . . .	17
2.1.3 Sampling Schemes . . . . .	17
2.1.4 Connection between Multinomial and Poisson distributions . . . . .	18
2.2 Loglinear Models in Two-Way Contingency Tables . . . . .	19
2.3 Likelihood in Loglinear Models . . . . .	20
2.4 Loglinear Association Parameters . . . . .	22

---

<b>3</b>	<b>Loglinear Model of Linear-by-Linear Association</b>	<b>23</b>
3.1	Association Models . . . . .	23
3.2	Bilinear Association Term and Ordinal Trends . . . . .	23
3.3	Choice of Scores and the Local Odds Ratios . . . . .	24
3.4	Linear-by-Linear Association Model Fitting and Inference . . .	26
3.5	Examples . . . . .	27
3.5.1	Astrology and Education . . . . .	27
3.5.2	Sexual Enjoyment Data . . . . .	30
<b>4</b>	<b>The Modified Signed Root of the Log-Likelihood Ratio</b>	<b>33</b>
4.1	The likelihoodAsy Package . . . . .	36
4.1.1	Examples . . . . .	37
<b>5</b>	<b>Simulation Study</b>	<b>51</b>
5.1	Conclusions . . . . .	84
	<b>Bibliography</b>	<b>85</b>
	<b>A R Code</b>	<b>87</b>

# Introduction

The objective of this project is to evaluate the behavior of likelihood-based inference in a model for the study of association in contingency tables.

The main disadvantages of the common loglinear models which allow association in contingency tables are that for ordinal classifications the models typically ignore ordinality and that for two-way contingency tables they are saturated. Instead, the presented model, the Linear-by-Linear Association model, describes association using just one additional parameter with respect to the independence model.

In particular, inference on the association parameter of the model is studied using two different statistics for significance testing. The first statistic is the usual signed root of the likelihood ratio, whereas the other one is a modification of the first. Both solutions are implemented by Bellio and Pierce (2016) in the R `likelihoodAsy`.

Since we deal with parametric statistical models and to do inference the likelihood is used, the first chapter presents an introduction to likelihood-based inference. Then the usual loglinear models used to model contingency tables are described in the second chapter, followed by the one which deals with the Linear-by-Linear Association model, in which two examples are introduced. In the fourth chapter the modification of the signed squared root of the likelihood ratio is described, as also its application using the R package.

Finally, the last chapter shows some simulations studies obtained both under association and independence and with several dimensions for the two-way table. The code used for the simulations, together with the parameters chosen for the simulation are listed in the Appendix.



# Chapter 1

## Likelihood

Fisher in the 1920s introduced inference procedures specifically focused on dealing with parametric statistical models.

Suppose we have a random variable  $Y$  and its observed value  $y$ , and that the probability density function of  $Y$  considering  $y$ ,  $f(y; \theta)$ , is known except for the value of a real parameter  $\theta$ . The objective is to make assertions about the distribution of  $Y$ , based on the observed data  $y$ .

### 1.1 Likelihood Function

The *likelihood function* provides the basis for procedures for point and interval estimate and for significance testing. In addition, thanks to the asymptotic theory, distribution problems linked to inference procedures based on the likelihood function find simple approximate solutions, even with moderate sample sizes.

The likelihood is defined to be a function of  $\theta$  for  $y$  fixed,  $L : \Theta \rightarrow \mathbb{R}^+$

$$L(\theta) = f(y; \theta), \quad \theta \in \Theta,$$

and when we want to underline the data dependence, the notation becomes  $L(\theta; y)$ .

Given the observed data  $y$ ,  $\theta_1 \in \Theta$  is more plausible than  $\theta_2 \in \Theta$  as value of the parameter of the data-generating model if  $L(\theta_1) > L(\theta_2)$ , and the two values can be compared through the ratio  $L(\theta_1)/L(\theta_2)$ . Therefore  $L(\theta)$

cannot be considered a density function on  $\Theta$ . Possible comparisons are the ones between pairs of parameter values, and concern with the adequacy of couples of plausible data-generating models. Since the ratio does not change if both  $L(\theta_1)$  and  $L(\theta_2)$  are multiplied by a constant, factors which do not depend on  $\theta$  can be ignored. Hence two likelihood functions which differ only by multiplicative constants are considered as equivalent.

Often it can be convenient to use the likelihood function on the logarithmic scale. We therefore define the *log-likelihood function*

$$l(\theta) = \log L(\theta),$$

with  $l(\theta) = -\infty$  if  $L(\theta) = 0$ . The monotonically increasing transformation is mathematically convenient and makes it simpler to describe the results. The comparison between log-likelihood functions is made through the difference, and if they differ in additive constants they are considered as equivalent.

## 1.2 Maximum Likelihood

Given the observed data, the *maximum likelihood estimate*  $\hat{\theta} \in \Theta$  is the value of  $\theta$  which maximizes the likelihood function, or equivalently the log-likelihood, since the idea is that the value of the likelihood function will be larger for values of  $\theta$  near the true value of  $\theta$ ,  $\theta_0$ , that is the one corresponding to the true data-generating model  $f^0(y)$ , i.e. such that  $f^0(y) = f(y; \theta_0)$ . Hence, we define

$$\hat{\theta} \in \Theta : L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta).$$

Generally we cannot say that  $\hat{\theta}$  exists and that it is unique. If  $\hat{\theta} = \hat{\theta}(y)$  exists and it is unique with probability one, then the random variable  $\hat{\theta} = \hat{\theta}(Y)$  is called *maximum likelihood estimator*.

Since many properties of interest of an estimator are asymptotic, it is sufficient for  $\hat{\theta}(y)$  to exist and to be unique with probability which tends towards one as  $n \rightarrow +\infty$  to define the maximum likelihood estimator  $\hat{\theta}(Y)$ .

## 1.3 Regularity Conditions

There are some regularity conditions that have to be satisfied in order to govern the behavior of the log-likelihood function and its derivatives, in a mathematically tractable manner. Models which satisfy these conditions are called *regular* models. The conditions are

- $\Theta$  is an open subset of  $\mathbb{R}^p$ ;
- the log-likelihood is a function differentiable at least three times, with continuous partial derivatives in  $\Theta$ ;
- the model is *identifiable*, i.e. different values of  $\theta$  correspond to different probability distributions. Formally,

$$\theta_1 \neq \theta_2 \Rightarrow \exists B : P(Y \in B; \theta_1) \neq P(Y \in B; \theta_2);$$

- the model is *correctly specified*, i.e. there is a value  $\theta_0 \in \Theta$  for which  $f(y; \theta_0) = f^0(y)$ ;
- the distribution or density functions have the same support, which does not depend on  $\theta$ .

More details can be found in the Section 3.2.3 of Azzalini (2001).

Dealing with regular parametric models,  $\hat{\theta}$  and the partial derivatives of the log-likelihood function with respect to the components of  $\theta$  are useful quantities for inference.

## 1.4 Score Function

The vector which contains the first order partial derivatives of the log-likelihood function

$$l_*(\theta) = \left( \frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right)$$

is called *score function*.

If the model is regular, the maximum likelihood estimate is the unique solution of the *likelihood equation*

$$l_*(\theta) = 0,$$

which is a system of equations when  $p > 1$ . Sometimes it is possible to solve the equation algebraically, but in general it is solved numerically.

## 1.5 Information

### 1.5.1 Observed Information

The matrix  $p \times p$  which contains the second order partial derivatives of the log-likelihood function, with changed sign,

$$j(\theta) = -l_{**}(\theta) = \left[ -\frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s} \right], \quad r, s = 1, \dots, p,$$

is called *observed information matrix*.

The observed information can be used to evaluate the accuracy in estimating the true parameter  $\theta_0$ , since the behavior of the log-likelihood function in a neighbourhood of  $\hat{\theta}$  is determined by the second order partial derivatives.

For  $p = 1$  the bigger  $j(\hat{\theta})$ , the more the likelihood is concentrated around  $\hat{\theta}$ , so  $j(\hat{\theta})$  is a measure of the information that  $y$  provides on  $\theta$ . It can be seen as an index of the speed of the fall of the log-likelihood moving away from  $\hat{\theta}$ , therefore of the degree of preference that the likelihood assign to  $\hat{\theta}$  compared with other values of  $\theta$ . When  $p > 1$ , it is a positive-definite matrix, and so the bigger the determinant, the smaller is the region of values of the parameter having the likelihood of the same order of magnitude of  $\hat{\theta}$ . In particular, if the model is regular, the log-likelihood function has approximately a parabolic shape, especially in the region with high likelihood.

### 1.5.2 Expected information

The *expected information* or *Fisher information* is the expected value of the observed information

$$i(\theta) = E_{\theta} \{j(\theta)\} = \left[ -E_{\theta} \left( \frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s} \right) \right], \quad r, s = 1, \dots, p.$$

## 1.6 Log-Likelihood Ratio Statistic

In a significance test, we compare the null hypothesis  $H_0 : \theta = \theta_0$  with the alternative hypothesis  $H_1 : \theta \neq \theta_0$ . A natural choice for such comparison is to use the *log-likelihood ratio statistic*

$$W(\theta_0) = 2 \left\{ l(\hat{\theta}) - l(\theta_0) \right\}.$$

Using a Taylor expansion, as shown in the Section 3.3 of Pace and Salvan (2001),  $l(\theta)$  in a neighbourhood of  $\hat{\theta}$  is approximately quadratic

$$l(\theta) \doteq l(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^2 j(\hat{\theta}).$$

Hence,  $l(\theta) - l(\hat{\theta}) \doteq -\frac{1}{2}(\theta - \hat{\theta})^2 j(\hat{\theta})$ . Since the log-likelihood ratio statistic is obtained calculating this quantity in  $\theta = \theta_0$  and multiplying it for  $-2$ , it is clear that high values of the statistic indicate that the null hypothesis has to be rejected.

Sometimes  $W(\theta_0)$  is a monotonically increasing function of a statistic  $t(y)$ , whose distribution under the null hypothesis is known, so it is possible to calculate the exact  $p$ -value

$$P_{\theta_0} \{W(\theta_0; Y) > W(\theta_0; y)\} = P_{\theta_0} \{t(Y) > t(y)\}.$$

More often, however, it is necessary to use an approximate null distribution. It can be shown that the asymptotic null distribution of  $W(\theta_0)$  is, under regularity conditions, chi-squared with  $p$  degrees of freedom.

## 1.7 Signed Root Log-likelihood Ratio Statistic

When  $p = 1$  other types of significance tests can be considered, such as test for  $H_0 : \theta = \theta_0$  versus the one-sided alternative hypothesis  $H_1 : \theta > \theta_0$ , or  $H_1 : \theta < \theta_0$ . In these cases it is natural to use the *signed root log-likelihood ratio statistic*

$$r(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0) \sqrt{W(\theta_0)},$$

where  $\text{sgn}$  is the sign function, whose value is 1 if its argument is positive, -1 if its argument is negative, 0 if its argument is 0.

As for  $W(\theta_0)$ , sometimes  $r(\theta_0)$  is a monotonically increasing function of a statistic whose distribution under the null hypothesis is known, so it is possible to calculate the exact  $p$ -value. However, usually, it is necessary to use an approximate null distribution. It is shown that the asymptotic null distribution of  $r(\theta_0)$  is, under regularity conditions,  $N(0, 1)$ .

It is clear that in the significance test having bilateral alternative hypothesis  $r(\theta_0)$  and  $W(\theta_0)$  lead to the same results.

## 1.8 Profile Likelihood

When  $p > 1$  we might be interested in only one component of the parameter, or on a subset of the parameter. So the parameter  $\theta$ , whose dimension is  $p$ , can be seen as partitioned into two components,  $\theta = (\tau, \zeta)$ . The first one is the *parameter of interest*  $\tau$ , whose dimension is  $k$ , with  $1 \leq k < p$ . The other one,  $\zeta$ , which has dimension  $p - k$ , is called *nuisance parameter*, since it is not of immediate interest.

The *profile likelihood* for  $\tau$  is the function

$$L_P(\tau) = L(\tau, \hat{\zeta}_\tau),$$

where  $\hat{\zeta}_\tau$  is the maximum likelihood estimate of  $\zeta$  in the submodel having  $\tau$  fixed. Moreover,  $l_P(\tau) = \log L_P(\tau)$  is called *profile log-likelihood function*.

For the null hypothesis  $H_0 : \tau = \tau_0$  it is natural to use the *profile log-likelihood ratio*

$$W_P(\tau_0) = 2 \left\{ l(\hat{\tau}, \hat{\zeta}) - l(\tau_0, \hat{\zeta}_0) \right\}.$$

The asymptotic null distribution of  $W_P(\tau_0)$  is, under regularity conditions, chi-squared with  $k$  degrees of freedom.

For the computation of  $W_P(\tau_0)$  it is sufficient to know the maximum likelihood estimate and the constrained maximum likelihood estimate calculated under the null hypothesis  $H_0 : \theta \in \Theta_0$ , where  $\Theta_0 \subset \Theta$

$$W_P^{H_0} = 2 \left\{ \sup_{\theta \in \Theta} l(\theta) - \sup_{\theta \in \Theta_0} l(\theta) \right\}.$$

When  $k = 1$ , it might be interesting also to consider significance tests such as  $H_0 : \tau = \tau_0$  versus the one-sided alternative  $H_1 : \tau > \tau_0$ , or  $H_1 : \tau < \tau_0$ .

In these cases we can use the signed root of  $W_P(\tau_0)$

$$r_P(\tau_0) = \text{sgn}(\hat{\tau} - \tau_0) \sqrt{W_P(\tau_0)}.$$

The asymptotic null distribution of  $r_P(\tau_0)$  is, under regularity conditions,  $N(0, 1)$ .



# Chapter 2

## Loglinear Models for Contingency Tables

### 2.1 Models for Count Data

There are a lot of applications in which the response variables have counts as their possible outcomes. Examples are the number of students who studied in the library of the Department of Statistical Sciences last week, or the number of ice-creams bought today in a given ice-cream shop of Padova between 3 p.m. and 4 p.m.

Also the frequencies of a frequency distribution are counts, and occur as entries in cells of contingency tables that cross-classify categorical variables such as, citing Agresti (2015), the number of people in a survey who are female, college educated, and agree that humans are responsible for climate change. In frequency distributions the total can be fixed or not, and in the first case the natural statistical model is the multinomial distribution. Nevertheless, since the multinomial model can be obtained by conditioning from a Poisson model, it is possible to treat in the same way contingency tables generated with different sampling schemes through the same loglinear Poisson models.

### 2.1.1 Poisson Model

The simplest statistical model for count data without any constraint on the total is the one in which the response variables, the counts, are realizations of independent random variables having Poisson distribution.

For counts of events that occur randomly over time or space the Poisson distribution is often utilized, in particular when outcomes in different time periods or regions are independent. Also, the Poisson model can be obtained as limit of the Binomial distribution  $Bi(m, \mu/m)$ , when  $m \rightarrow \infty$  and  $\pi = \mu/m$  is very small.

When there are also concomitant variables, we are interested in models which explain the dependence of the means  $\mu_i$  by means of the values of the explanatory variables. The null model is the one in which the means  $\mu_i$  are constant.

The Poisson Generalized Linear Model assumes  $Y_i \sim Poisson(\mu_i)$  and  $g(\mu_i) = \eta_i = x_i\beta$ , where  $x_i$  is a vector of explanatory variables. The likelihood function is

$$l(\beta) = \sum_{i=1}^n \{y_i \log(\mu_i) - \mu_i\}, \quad (2.1)$$

with  $\mu_i = g^{-1}(x_i\beta)$ . The likelihood equations for  $\beta$  are

$$l_* = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ir}}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad r = 1, \dots, p.$$

The canonical link for a Poisson generalized linear model is the log link  $\log(\mu_i) = \eta_i$ . The model then is called *Poisson loglinear model* and implies  $Y_i \sim P(e^{\mu_i})$ , where the means satisfy the exponential relation

$$\mu_i = \exp \{x_i^T \beta\}.$$

A 1-unit increase in  $x_{ir}$  has a multiplicative impact on  $\mu_i$  of a factor  $e^{\beta_r}$ .

Except for very simple cases, the maximum likelihood estimate is obtained numerically through the Iteratively Reweighted Least Squares algorithm, which is implemented in most statistical softwares. The estimated covariance matrix of  $\hat{\beta}$  is

$$\text{var}(\hat{\beta}) = (X^T \hat{W} X)^{-1},$$

where, with the log link,  $W$  is a diagonal matrix with elements  $\hat{\mu}_1, \dots, \hat{\mu}_n$ .

The deviance of a Poisson generalized linear model is

$$D(y, \hat{\mu}) = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right].$$

It can be utilized to test the goodness of fit of the model when the number  $n$  of Poisson observation is fixed and the means  $\mu_i$  diverge. This happens, for instance, in contingency tables having a fixed number of cells and large sample size.

Also the residuals can provide useful information, underlining the presence of unusual observation or patterns. Another possible way to judge goodness of fit is to compare the sample proportion of  $\{0, 1, 2, \dots\}$  observation with the fitted frequencies  $\sum_{i=1}^n Pr(Y_i = j; \hat{\beta})/n$ .

### 2.1.2 Loglinear Models for Contingency Tables

Binomial and multinomial models are often used with categorical response variables. With multivariate categorical responses they can be applied to each marginal distribution, or alternatively the multinomial model can be used to analyze the joint distribution, to investigate if the data present particular independence, association or interaction schemes.

Generally, inference on the parameters of the multinomial models coincides with inference based on the Poisson models, in particular for independent Poisson counts in the cells of a contingency table. The equivalence occurs because the multinomial model can be obtained from a Poisson model through conditioning on the sample size, or on the rows or column totals, as shown in Section 2.1.4.

### 2.1.3 Sampling Schemes

Counts can be generated from different sampling schemes depending on the fact that the sample size, or the marginal totals for a variable, are fixed or not.

If there are no constraints on the totals, a simple model assumes  $y_{ij}$  realization of  $Y_{ij} \sim P(\mu_{ij})$ , and it is supposed that every  $y_{ij}$  is independent

from the other observations. For a two-way table with  $r$  rows and  $c$  columns, the likelihood function is

$$\prod_{i=1}^r \prod_{j=1}^c \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}.$$

If instead the total count, that is the sample size  $s = \sum_{i=1}^r \sum_{j=1}^c y_{ij}$ , is fixed, then the distribution is a multinomial and the likelihood function is

$$\frac{s!}{\prod_{i,j} y_{ij}!} \prod_{i,j} \pi_{ij}^{y_{ij}},$$

with  $0 < \pi_{ij} < 1$  and  $\sum_{i,j} \pi_{ij} = 1$ . The parameters  $\pi_{ij}$  and the  $\mu_{ij}$  are linked through the relation

$$\pi_{ij} = \frac{\mu_{ij}}{\sum_{i,j} \mu_{ij}}.$$

A contingency table can also be obtained with sampling schemes in which the row totals - or column totals - are fixed. The statistical model then will be the product of the multinomial distributions related to each row (or column).

### 2.1.4 Connection between Multinomial and Poisson distributions

It is assumed that  $(y_1, \dots, y_n)$  are realizations of independent Poisson random variables with means  $(\mu_1, \dots, \mu_n)$ , in particular  $\mu_i = e^{x_i^T \beta}$  in Poisson models having canonical link. Also the total  $s = \sum_i y_i$  has a Poisson distribution, with mean  $\mu = \sum_i \mu_i$ .

Conditioning on  $s$ ,  $(Y_1, \dots, Y_n)$  will have no longer Poisson distribution because they are no longer independent, since the sample size cannot exceed  $s$  and the value of one influences the possible range for the others.

The conditional probability is

$$\begin{aligned} P \left\{ (Y_1 = y_1, \dots, Y_n = y_n) \mid \sum_{j=1}^n Y_j = s \right\} &= \frac{P(Y_1 = y_1, \dots, Y_n = y_n)}{P(\sum_j Y_j = s)} \\ &= \frac{\prod_i e^{-\mu_i} \mu_i^{y_i} / y_i!}{e^{-\sum_j \mu_j} (\sum_j \mu_j)^s / s!} = \frac{s!}{\prod_i y_i!} \prod_{i=1}^n \pi_i^{y_i}, \end{aligned} \quad (2.2)$$

where  $\pi_i = \mu_i / \sum_j \mu_j$ . The result is the multinomial distribution characterized by sample size  $s$  and probabilities  $\pi_i$ .

## 2.2 Loglinear Models in Two-Way Contingency Tables

Suppose we have  $Y_{ij}$  independent counts having Poisson distributions and means  $\mu_{ij}$  satisfying

$$\mu_{ij} = \mu\phi_i\psi_j,$$

with  $\phi_i > 0$ ,  $\psi_j > 0$  and  $\sum_i \phi_i = \sum_j \psi_j = 1$ .

Under independence, the mean of the Poisson loglinear model  $\mu_{ij}$  has a loglinear structure

$$\log \mu_{ij} = \beta_0 + \beta_i^X + \beta_j^Z, \quad (2.3)$$

where  $\beta_i^X$  and  $\beta_j^Z$  are the effects of the two classification variables  $X$  and  $Z$ . Identifiability requires a constraint on  $\{\beta_i^X\}$  and one on  $\{\beta_j^Z\}$ .

As shown in (2.2), conditioning on the total  $s$ , the cells counts have a multinomial distribution and probabilities

$$\pi_{ij} = \frac{\mu_{ij}}{\sum_{i,j} \mu_{ij}} = \frac{\mu_{jh}}{\mu} = \phi_i\psi_j.$$

Since  $\sum_i \phi_i = \sum_j \psi_j = 1$ , we have  $\phi_i = \pi_{i+}$  and  $\psi_j = \pi_{+j}$ . Hence, under independence,  $\pi_{ij} = \pi_{i+}\pi_{+j}$ .

Exponentiating the Poisson loglinear model (2.3) and dividing both sides by  $\mu$ , the intercept parameter  $\beta_0$  cancels. Therefore, the Poisson model has  $1 + (r-1) + (c-1)$  parameters, whereas the multinomial one  $(r-1) + (c-1)$ .

The model matrix for the Poisson loglinear model contains the values of the indicator variables for the two categorical variables. For instance, for  $r = c = 2$  and  $\beta_1^X = \beta_1^Z = 0$ , the model is

$$\log \mu = \begin{bmatrix} \log \mu_{11} \\ \log \mu_{12} \\ \log \mu_{21} \\ \log \mu_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_2^X \\ \beta_2^Z \end{bmatrix}.$$

The log-likelihood function of the model is

$$\begin{aligned} l(\beta_0, \beta^X, \beta^Z) &= \sum_i \sum_j y_{ij} \log \mu_{ij} - \sum_i \sum_j \mu_{ij} \\ &= \beta_0 y_{++} + \sum_i \beta_i^X y_{i+} + \sum_j \beta_j^Z y_{+j} - \sum_i \sum_j \exp(\beta_0 + \beta_i^X + \beta_j^Z), \end{aligned}$$

where  $y_{++} = \sum_{i,j} y_{ij}$ ,  $y_{i+} = \sum_j y_{ij}$  and  $y_{+j} = \sum_i y_{ij}$ .

The likelihood equations impose the identity

$$\sum_i y_i x_{ir} = \sum_i \mu_i x_{ir} \quad (2.4)$$

which is satisfied for  $\hat{\mu}_{++} = \sum_{i,j} \hat{\mu}_{ij} = y_{++}$ ,  $\hat{\mu}_{i+} = \sum_j \hat{\mu}_{ij} = y_{i+}$  and  $\hat{\mu}_{+j} = \sum_i \hat{\mu}_{ij} = y_{+j}$ . The maximum likelihood estimates then satisfy  $\hat{\mu}_{ij} = y_{i+}y_{+j}/s$ , where  $s = y_{++}$ .

Conditioning on  $s$  and maximizing the log-likelihood of the multinomial distribution  $\sum_i \sum_j y_{ij} \log \pi_{ij}$ , we obtained the same fit and  $\pi_{ij}$  is

$$\pi_{ij} = \frac{\mu_{ij}}{\sum_i \sum_j \mu_{ij}} = \frac{e^{\beta_i^X + \beta_j^Z}}{\sum_i \sum_j e^{\beta_i^X + \beta_j^Z}},$$

so  $\hat{\pi}_{ij}$  is the product of the sample marginal proportion

$$\hat{\pi}_{ij} = \frac{y_{i+}}{s} \frac{y_{+j}}{s}.$$

To test the goodness of fit of the independence model the Pearson statistic is used

$$X^2 = \sum_i \sum_j \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}.$$

The null approximate distribution is chi-squared with  $(r-1)(c-1)$  degrees of freedom, for both Poisson and multinomial models.

## 2.3 Likelihood in Loglinear Models

In this paragraph we show the connection between the Poisson likelihood function and the multinomial one, through the comparison of the log-likelihood functions.

As mentioned above, the Poisson log-likelihood function is

$$l^P(\mu) = \sum_i (y_i \log \mu_i - \mu_i)$$

and the multinomial one is

$$l^M(\mu) = \sum_i (y_i \log \pi_i) = \sum_i (y_i \log \mu_i - y_i)$$

since  $\mu_i = s\pi_i$ . Hence the two functions differ by an additive constant under the condition for which the model includes a constant term so that

$\sum_i \hat{\mu}_i = \sum_i y_i$  for the identity imposed by the likelihood equations of the Poisson model (2.4).

Therefore it is possible to estimate loglinear models having multinomial distribution through the Poisson likelihood function. In fact, the Poisson loglinear model (2.3) can be written highlighting the parameter  $\gamma$  related to the constant explanatory variable

$$\log \mu_i = \gamma + z_i^T \beta$$

where  $z_i^T \beta$  is the component connected to the other explanatory variables  $z_i$ , in which  $\beta$  has one parameter less than usual. So  $l^P(\mu)$  becomes

$$\begin{aligned} l^P(\mu) &= \sum_i y_i \log \mu_i - \sum_i \mu_i \\ &= \sum_i y_i (\gamma + z_i^T \beta) - \sum_i \mu_i \\ &= s\gamma + \left( \sum_i y_i z_i^T \right) \beta - \sum_i \mu_i \end{aligned}$$

and given that  $\sum_i \mu_i = \mu_{++} = \sum_i \exp(\gamma + z_i^T \beta)$  we have  $\log \mu_{++} = \gamma + \log \left\{ \sum_i \exp(z_i^T \beta) \right\}$ , and  $\pi_i = \mu_i / \sum_{h=1}^{r-1} \mu_h = \exp(z_i^T \beta) / \sum_{h=1}^{r-1} \exp(z_h^T \beta)$ ,  $l^P(\mu)$  can be split as follows

$$\begin{aligned} l^P(\mu_{++}, \beta) &= \left\{ \left( \sum_i y_i z_i^T \right) \beta - s \log \left[ \sum_i \exp(z_i^T \beta) \right] \right\} + [s \log \mu_{++} - \mu_{++}] \\ &= l^M(\beta) + [s \log \mu_{++} - \mu_{++}]. \end{aligned}$$

Since the second term on the right-hand side of the above equation does not depend on the parameter  $\beta$ , the value  $\hat{\beta}$  which maximizes  $l^P$  is the same  $\hat{\beta}$  which maximizes  $l^M$ . Also the second order partial derivatives with respect to  $\beta$  are the same, so also the standard errors will be the same. More details can be found in the Section 6.4.3 of Azzalini (2001).

## 2.4 Loglinear Association Parameters

The association between the two classification variables can be permitted modifying the loglinear model (2.3) adding the interaction terms

$$\log \mu_{ij} = \beta_0 + \beta_i^X + \beta_j^Z + \beta_{ij}^{XZ}. \quad (2.5)$$

The interaction parameters can be specified so that they equal the cross-product of the of the indicator variables of  $X$  and  $Z$ . After stating appropriate constraints for identifiability, such as  $\beta_{1j}^{XZ} = \beta_{i1}^{XZ} = 0$  for all  $i$  and  $j$ , the number of the interaction terms is  $(r-1)(c-1)$ , so (2.5) corresponds to the saturated model.

Interaction parameters are related to log odds ratios. For instance, for  $r = c = 2$ , the log odds ratio is

$$\begin{aligned} \log \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} &= \log \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} = \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21} = \\ &= (\beta_0 + \beta_1^X + \beta_1^Z + \beta_{11}^{XZ}) + (\beta_0 + \beta_2^X + \beta_2^Z + \beta_{22}^{XZ}) + \\ &\quad - (\beta_0 + \beta_1^X + \beta_2^Z + \beta_{12}^{XZ}) - (\beta_0 + \beta_2^X + \beta_1^Z + \beta_{21}^{XZ}) \\ &= \beta_{11}^{XZ} + \beta_{22}^{XZ} - \beta_{12}^{XZ} - \beta_{21}^{XZ} \end{aligned}$$

Under the identifiability constraints, the log odds ratio simplifies to  $\beta_{22}^{XZ}$ .

# Chapter 3

## Loglinear Model of Linear-by-Linear Association

### 3.1 Association Models

The main disadvantages of loglinear models that allow association are that for ordinal classifications the models ignore ordinality and that for two-way contingency tables they are saturated.

Association models have the advantage that the association is described by a simple structure, and that the models can be tested.

Here we consider a two-way contingency table in which both variables are ordinal.

### 3.2 Bilinear Association Term and Ordinal Trends

The bilinear association model assigns ordered scores to both the row ( $u_1 \leq u_2 \leq \dots \leq u_r$ ) and the column categories ( $v_1 \leq v_2 \leq \dots \leq v_c$ ).

The Linear-by-Linear Association model (L x L) is defined as

$$\log \mu_{ij} = \beta_0 + \beta_i^X + \beta_j^Z + \tau u_i v_j, \quad (3.1)$$

with constraints such as  $\beta_r^X = \beta_c^Z = 0$  for identifiability. It is the special case of the saturated model in which the association term has the form  $\beta_{ij}^{XZ} =$

$\tau u_i v_j$ . Using just one more parameter than the independence model, this simple model can describe the association between the variables, while the general model uses  $(r - 1)(c - 1)$  parameters.

When the variables are independent, the model turns into the special case in which  $\tau = 0$ . The deviation of  $\mu_{ij}$  from the independence pattern is represented by  $\exp(\tau u_i v_j)$ . In particular  $\tau$  highlights the direction and the power of the association, so that when  $\tau > 0$   $Z$  tends to increase as  $X$  increases, hence  $\tau$  underlines the presence of a positive relation between the variables. When  $\tau < 0$ ,  $Z$  tends to decrease as  $X$  increases, pointing out a negative relationship between the two. When there is a positive or a negative association between the variables, the model fits better than the independence one.

At a given level of  $X$  the deviation of  $\log \mu_{ij}$  from independence is linear in the  $Z$  scores  $\{v_j\}$  and vice versa. This property is the reason of the name of the model (3.1): *Linear-by-Linear Association model*. For instance, the deviation in a fixed column  $j$  is a linear function of  $X$ , calculated multiplying  $\tau v_j$  for the score of the row.

The model entails that the greatest departures from independence are in the four corners of the table, and that, under independence, expected frequencies are larger than expected in cells where  $X$  and  $Z$  are both high or both low.

In  $2 \times 2$  tables, the log odds ratio for row categories is

$$\log \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} = \tau(u_2 - u_1)(v_2 - v_1)$$

The log odds ratio is stronger as  $|\tau|$  increases and for pairs of categories that are farther apart.

### 3.3 Choice of Scores and the Local Odds Ratios

A disadvantage of the model is that the interpretation of  $\tau$  depends on the choice of the scores. When the row scores  $u_2 - u_1 = \dots = u_r - u_{r-1}$  and the column ones  $v_2 - v_1 = \dots = v_c - v_{c-1}$  are equidistant the interpretation is simple, and the *local odds ratios* for adjacent rows and adjacent columns

have the same value. Duncan and Goodman (1979) called this case *uniform association*. For unit-spaced scores the value of the local odds ratio is  $e^\tau$ .

When the classification variables are quantitative and for which values have been grouped into ordered categories, it is reasonable to choose scores that approximate distances between the midpoints of categories, so that  $e^\tau$  represents the odds ratio for unit distances. Anyway it is not necessary to choose scores that approximate distances between categories or reasonable scalings of ordinal variables, since the scores only imply a pattern for the local odds ratios. When the model fits well with equally spaced scores, the uniform local odds ratio describes the association in any case, both the scores are reasonable indexes and true distances between categories.

Sets of scores having the same spacings between the levels of the variables lead to the same maximum likelihood estimate  $\hat{\tau}$ , fitted values  $\{\hat{\mu}_{ij}\}$  and values  $X^2, G^2$  to test the goodness of fit of the model. A linear transformation of the scores does not affect the inferential results or the goodness of fit. For instance, two sets of scores with one being a linear transformation of the other lead to the same fitted values, but to a rescaled  $\hat{\tau}$ .

It is sometimes useful to standardize the scores, so that they satisfy

$$\begin{aligned}\sum u_i \pi_{i+} &= \sum v_j \pi_{+j} = 0 \\ \sum u_i^2 \pi_{i+} &= \sum v_j^2 \pi_{+j} = 1.\end{aligned}$$

Then  $\tau$  represents the log odds ratio for a standard deviation increase in each variable. Given the standard deviations for the original scores  $(\sigma_X, \sigma_Z)$ , the effect for the original scores  $\tau$  and the effect for the standardized scores  $\tau^*$ ,  $\tau^*$  equals the product of  $\tau$  and the standard deviations

$$\tau^* = \tau \sigma_X \sigma_Z.$$

The uniform association model fits well when the underlying distribution is approximately bivariate normal, then  $\tau^*$  approximates  $\rho^* / \{1 - (\rho^*)^2\}$ , where  $\rho^*$  is the correlation between the underlying continuous variables.

### 3.4 Linear-by-Linear Association Model Fitting and Inference

In a two-way contingency table, for  $y_{ij}$  realizations of independent Poisson random variables, the log-likelihood function is

$$l(\mu) = \sum_i \sum_j y_{ij} \log \mu_{ij} - \sum_i \sum_j \mu_{ij}$$

For the Linear-by-Linear Association model (3.1), the log-likelihood function becomes

$$\begin{aligned} l(\mu) &= s\beta_0 + \sum_i y_{i+}\beta_i^X + \sum_j y_{+j}\beta_j^Z + \tau \sum_i \sum_j u_i v_j y_{ij} \\ &\quad - \sum_i \sum_j \exp(\beta_0 + \beta_i^X + \beta_j^Z + \tau u_i v_j). \end{aligned}$$

The likelihood equations of the model lead to

$$\begin{aligned} \hat{\mu}_{i+} &= y_{i+}, & i &= 1, \dots, r, \\ \hat{\mu}_{+j} &= y_{+j}, & j &= 1, \dots, c, \\ \sum_i \sum_j u_i v_j \hat{\mu}_{ij} &= \sum_i \sum_j u_i v_j y_{ij}. \end{aligned}$$

Let  $p_{ij} = y_{ij}/s$  and  $\hat{\pi}_{ij} = \hat{\mu}_{ij}/s$ . The likelihood equations imply that  $\hat{\pi}_{i+} = p_{i+}$  and  $\hat{\pi}_{+j} = p_{+j}$  for all  $i$  and  $j$ , and  $\sum_i \sum_j u_i v_j \hat{\pi}_{ij} = \sum_i \sum_j u_i v_j p_{ij}$ . Therefore the marginal distributions are identical for the fitted and observed distributions, so that also the correlation  $\hat{\rho}$  between the scores for  $X$  and  $Z$  is the same for both distributions. The larger  $|\hat{\rho}|$  is, the more the data are far from independence.

Since the scores are fixed, the Linear-by-Linear Association model (3.1) has only one more parameter than the independence model,  $\tau$ . Its residual degrees of freedom are

$$(r-1)(c-1) - 1 = rc - r - c,$$

so the model is unsaturated except for  $2 \times 2$  tables. The Pearson  $X^2$  and deviance  $G^2$  statistics are used to test the goodness of fit of the model and

have, for large samples, approximate null chi-squared distribution with  $rc - r - c$  degrees of freedom.

Given the deviance statistic for testing the goodness of fit of the independence model  $G^2(\text{I})$  and the deviance for testing the one of the Linear-by-Linear Association model  $G^2(\text{L x L})$ , to detect positive or negative trends we use the likelihood ratio statistic computed under the null hypothesis  $H_0 : \tau = 0$ , which equals

$$G^2(\text{I} \mid \text{L x L}) = G^2(\text{I}) - G^2(\text{L x L})$$

The asymptotic null distribution is chi-squared with 1 degree of freedom.

Other possibilities are the Wald statistic and the score statistic. When the data are highly sparse, another possibility is a bootstrap test using  $G^2(\text{I} \mid \text{L x L})$  for samples obtained from a multinomial distribution having probabilities based on the fit of the independence model.

Because of the noncentrality of the chi-squared asymptotic distribution, the power of a chi-squared test increases when the degrees of freedom decrease. When the model holds,  $G^2(\text{I} \mid \text{L x L})$  is asymptotically more powerful than  $G^2(\text{I})$ , since even if the noncentrality is the same for both,  $G^2(\text{I} \mid \text{L x L})$  has 1 degree of freedom compared to  $(r - 1)(c - 1)$  for  $G^2(\text{I})$ , so that as  $r$  and  $c$  increase, the noncentrality remains focused on 1 degree of freedom for  $G^2(\text{I} \mid \text{L x L})$ , but the degrees of freedom increase for  $G^2(\text{I})$ .

## 3.5 Examples

### 3.5.1 Astrology and Education

This example is taken from the Section 6.2.5 of Agresti (2010), and deal with the connection between education and belief about Astrology. The two classification variables indicate the highest degree of the person and the opinion the person has regarding the expression "Astrology is scientific". Table 3.1 shows the observed data, whose sample size is 1793.

The output for fitting the Linear-by-Linear Association model is obtained through the function `glm`, indicating `family = poisson` and passing

the counts as response and the two classification variables and the product of the scores as explanatory variables as follows

```
LxL <- glm(y ~ astro1 + degree1 + uv, family = poisson)
summary(LxL)
```

**Table 3.1:** Education and Belief About Astrology

Highest Degree	Astrology is Scientific		
	Not at all	Sort of	Very
< High School	98	84	23
High School	574	286	50
Junior College	122	44	4
Bachelor	268	57	11
Graduate	148	23	1

Using unit-spaced scores, the code above gives the output

```
Call:
glm(formula = y ~ astro1 + degree1 + uv, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.69382  -0.48694  -0.04418   0.24369   2.08966

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3153     0.4948   0.637 0.523971
astro11     -1.5585     0.1453 -10.725 < 2e-16 ***
astro12     -0.2885     0.0821  -3.515 0.000441 ***
degree11     4.3349     0.4672   9.278 < 2e-16 ***
degree12     4.8562     0.3656  13.283 < 2e-16 ***
degree13     2.1538     0.2672   8.061 7.57e-16 ***
degree14     1.7679     0.1559  11.339 < 2e-16 ***
uv           0.3898     0.0421   9.260 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for `poisson family` taken to be 1)

`Null deviance:` 2202.6574 on 14 degrees of freedom  
`Residual deviance:` 6.8389 on 7 degrees of freedom  
`AIC:` 108.87

Number of Fisher Scoring iterations: 4

Table 3.2 shows the fitted values for the model using equally spaced scores.

**Table 3.2:** Education and Belief About Astrology with Fit of Linear-by-Linear Association Model

Highest Degree	Astrology is Scientific		
	Not at all	Sort of	Very
< High School	104.6	78.4	22.0
High School	567.3	287.9	54.8
Junior College	122.5	42.1	5.4
Bachelor	268.2	62.4	5.4
Graduate	147.4	23.2	1.4

The model fits very well, since it has a residual deviance of 6.8 with 7 degrees of freedom, especially compared to the independence model whose results regarding the goodness of fit are substantially worse:

`Null deviance:` 2202.66 on 14 degrees of freedom  
`Residual deviance:` 107.61 on 8 degrees of freedom  
`AIC:` 207.64

Testing the null hypothesis  $H_0 : \tau = 0$ , the likelihood-ratio statistic equals  $G^2(I|L \times L) = 107.6 - 6.8 = 100.8$ , showing strong evidence of association.

For the scores of the columns  $v$  the reversing category order of the variable "Astrology is Scientific" is used, so that "Not at all" corresponds to 3 and "Very" to 1. Using this codification the estimate of the parameter  $\tau$  is 0.3898, showing that subjects having lower levels of education tend to see

more scientific basis to Astrology, whereas it would be negative and equal to  $-0.3898$ . The estimated common local odds ratio is  $\exp(0.3898) = 1.46$ .

The model predicts the greatest departures from independence in the four corners of the table, where the odds ratio equals

$$\exp[\hat{\tau}(u_5 - u_1)(v_3 - v_1)] = \exp[0.3898(5 - 1)(3 - 1)] = 22.61,$$

so that the estimated odds of believing that Astrology is very scientific versus not scientific at all for those having less than an high school degree are 22.6 times the odds for those with a graduate degree. The 95% Wald confidence interval for the common local odds ratio is  $\exp[0.3898 \pm (1.96)(0.042)] = (1.36, 1.60)$ , and the corresponding 95% Wald confidence interval for the odds ratio for the four corner cells is  $\exp[8(0.3898) \pm 8(1.96)(0.042)] = (11.7, 43.7)$ .

Using standardized scores, the parameter estimate results

$$\hat{\tau}^* = 0.3898(1.464)(0.8452) = 0.4823.$$

Solving the equation  $\hat{\tau}^* = \hat{\rho}^*/[1 - (\hat{\rho}^*)^2]$  for  $\hat{\rho}^*$ , we obtain  $\hat{\rho}^* = 0.4037$ , so that if the underlying distribution is approximately bivariate normal the estimate of the correlation is 0.4037, which is a mild positive association.

### 3.5.2 Sexual Enjoyment Data

This example is taken from the Section 3.1 of Kolassa and Tanner (1994), and deals with the sexual enjoyment of 91 married couples. The data are shown in Table 3.3.

**Table 3.3:** Sexual Enjoyment Data

Husband's response	Wife's response			
	Never or occasionally	Fairly often	Very often	Almost always
Never or occasionally	7	7	2	3
Fairly often	2	8	3	7
Very often	1	5	4	9
Almost always	2	8	9	14

The output for fitting the Linear-by-Linear Association model using equally spaced scores is

```
LxL <- glm(y ~ wife + husb + uv, family = poisson)
summary(LxL)

Call:
glm(formula = y ~ wife + husb + uv, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.80867  -0.53562  -0.07395   0.34885   0.93544

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.28172    0.33474   3.829 0.000129 ***
wife2        0.19717    0.38813   0.508 0.611451
wife3       -1.00343    0.56257  -1.784 0.074479 .
wife4       -1.26029    0.80242  -1.571 0.116276
husb2       -0.64346    0.37700  -1.707 0.087861 .
husb3       -1.48203    0.55779  -2.657 0.007885 **
husb4       -1.80654    0.81548  -2.215 0.026738 *
uv           0.28851    0.09451   3.053 0.002269 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 33.5846  on 15  degrees of freedom
Residual deviance:  5.0051  on  8  degrees of freedom
AIC: 75.471

Number of Fisher Scoring iterations: 4
```

Table 3.4 shows the fitted values for the model.

The model has a residual deviance of 5 with 8 degrees of freedom, whereas the goodness of fit of the independence model is

**Table 3.4:** Sexual Enjoyment Data with Fit of Linear-by-Linear Association Model

Husband's response	Wife's response			
	Never or occasionally	Fairly often	Very often	Almost always
Never/occasionally	4.8	7.8	3.1	3.2
Fairly often	3.4	7.3	3.9	5.4
Very often	1.9	5.6	4.0	7.4
Almost always	1.9	7.2	6.9	17.0

**Null deviance:** 33.585 on 15 degrees of freedom

**Residual deviance:** 15.486 on 9 degrees of freedom

**AIC:** 83.952

Testing the null hypothesis  $H_0 : \tau = 0$ , the likelihood-ratio statistic equals  $G^2(I|L \times L) = 15.486 - 5.005 = 10.48$ , showing evidence of association assuming a significance level of 0.05 ( $p$ -value = 0.0012).

The estimated common local odds ratio is  $\exp(0.2885) = 1.33$ , whereas in the four corners of the table the odds ratio equals

$$\exp[\hat{\tau}(u_4 - u_1)(v_4 - v_1)] = \exp[0.2885(4 - 1)(4 - 1)] = 13.42.$$

The 95% Wald confidence interval for the common local odds ratio is  $\exp[0.2885 \pm (1.96)(0.0945)] = (1.11, 1.61)$ , and the 95% Wald confidence interval for the odds ratio for the four corner cells is  $\exp[9(0.2885) \pm 9(1.96)(0.0945)] = (2.53, 71.1)$ .

Using standardized scores instead the parameter estimate results

$$\hat{\tau}^* = 0.2885(1.155)(1.155) = 0.3849.$$

Solving the equation  $\hat{\tau}^* = \hat{\rho}^*/[1 - (\hat{\rho}^*)^2]$  for  $\hat{\rho}^*$ , we obtain  $\hat{\rho}^* = 0.34$ , so that if the underlying distribution is approximately bivariate normal the estimate of the correlation is 0.34, which is a relatively weak positive association.

## Chapter 4

# The Modified Signed Root of the Log-Likelihood Ratio

In this chapter modern higher-order likelihood inference is briefly reviewed. In particular a special likelihood-based procedure which modifies the usual signed root of the log-likelihood ratio for higher accuracy is discussed.

As described in the Section 1.8, the signed root of  $W_P(\tau_0)$  has to  $O(n^{-1/2})$ , i.e. to first order, an asymptotic null distribution  $N(0, 1)$  under regularity conditions. Then for observed data  $y$ ,

$$Pr \{r_P(Y, \tau_0) \leq r_P(y, \tau_0); \tau(\theta) = \tau_0\} = \Phi \{r_P(y, \tau_0)\} \{1 + O(n^{-1/2})\},$$

where  $\Phi$  is the standard normal distribution function and  $n$  is the sample size.

A modification of  $r_P(\tau_0)$ , commonly denoted by  $r^*(\tau_0)$ , has an accuracy that can be formalized as the second-order result

$$Pr \{r_P(Y, \tau_0) \leq r_P(y, \tau_0); \tau(\theta) = \tau_0\} = \Phi \{r_P^*(y, \tau_0)\} \{1 + O(n^{-1})\}. \quad (4.1)$$

The quantity  $r^*(\tau_0)$  was derived by Barndorff-Nielsen in a form generally difficult to compute. After that a lot of approximations came to notice, and the one utilized in the accompanying software is the version developed by Skovgaard (1996) Bernoulli.

At first the distribution of  $\hat{\theta}$  is estimated using the likelihood ratio, and

then, through transformation and integration, we obtain an accurate approximation to the distribution of  $r(\tau_0)$ .

In general  $\hat{\theta}$  is required to be a sufficient statistic, as indeed happens in log-linear models. When it is not, an approximate *ancillar statistic*  $a$  is used so that, conditioning on it,  $\hat{\theta}$  is conditionally sufficient. An ancillar statistic brings information about the accuracy of  $\hat{\theta}$ , but not the value of  $\theta$  itself. Anyway, despite conditioning on the ancillar statistic is important in considering the distribution of  $\hat{\theta}$ , it is not that much for the distribution of  $r(\tau_0)$ , since the likelihood ratio is to second order stochastically independent of any ancillary.

The ancillary information is based on the ratio of observed to expected information  $i(\hat{\theta})^{-1}j(\hat{\theta})$ , and in order to make the distribution of the ancillary more constant in  $\theta$ , it can be rescaled dividing it by  $n^{1/2}$  times its estimated asymptotic standard deviation, obtaining an ancillar statistic denoted by  $a$ , whose dimension does not depend on the sample size. Skovgaard (1996) established that this statistic is locally second order ancillary, so that the distribution of  $a$  depends on  $\theta$  only in terms of  $O(n^{-1})$ , for  $\theta$ -variations of  $O(n^{-1/2})$ . Given  $a$ , the accuracy (4.1) becomes

$$Pr \{r_P(Y) \leq r_P(y)|a\} = \Phi \{r_P^*(y)\} \{1 + O(n^{-1})\},$$

for  $\|\theta - \hat{\theta}\| = O(n^{-1/2})$ , but the result is valid also without conditioning on the ancillary  $a$ , since likelihood ratio statistics are to second order independent of any ancillary.

The likelihood ratio approximation to the density of  $\hat{\theta}$ , often called Barnodorff-Nielsen's  $p^*$  formula, when  $\hat{\theta}$  is a sufficient statistic, with  $p \geq 1$ , is

$$\begin{aligned} p^*(\hat{\theta}; \theta) &= \frac{|j(\hat{\theta})|^{1/2} L(\theta)}{(2\pi)^{p/2} L(\hat{\theta})} \\ &= L(\hat{\theta}) \{1 + O(n^{-1})\} \end{aligned} \tag{4.2}$$

For the case in which  $\hat{\theta}$  is not sufficient, it is necessary an ancillar statistic  $a$  so that the quantity  $(\hat{\theta}, a)$  is sufficient and  $\hat{\theta}$  is sufficient for the conditional distribution of  $\hat{\theta}|a$ .

The same approximation formula as (4.2) used to approximate the density

conditional on  $a$ , is

$$\begin{aligned} p^*(\hat{\theta}|a) &= \frac{|j(\hat{\theta})|^{1/2} L(y, \theta)}{(2\pi)^{p/2} L(y, \hat{\theta})} \\ &= L(\hat{\theta}|a) \{1 + O(n^{-1})\} \end{aligned}$$

which is often referred to as "Barndorff-Nielsen's *magic formula*".

When  $p = 1$ , the distribution of  $r$  has density, under regularity conditions,

$$p^*(r|a, \theta) = \left| \frac{\partial r}{\partial \hat{\theta}} \right|^{-1} p^*(\hat{\theta}|a, \theta).$$

Since this is not convenient to use, a second-order approximation is used. From differentiating  $r^2$ , we obtain that  $\partial r / \partial \hat{\theta} = \left[ \partial \{l(\hat{\theta}) - l(\theta)\} / \partial \hat{\theta} \right] / r$ . After denoting  $\partial \{l(\hat{\theta}) - l(\theta)\} / \partial \hat{\theta}$  by  $u$ ,  $p^*$  can be rewritten as follows

$$p^*(r|a) = \left| \frac{u}{r} \right|^{-1} (2\pi)^{-1/2} \exp(-r^2/2).$$

After raising  $|u/r|$  to the exponential, completing the square and dropping the second-order term  $(u/r)^2$ ,  $p^*$  becomes

$$p^*(r|a) = (2\pi)^{-1/2} \exp(-(r^*)^2/2) \quad (4.3)$$

where  $r^* = r + r^{-1} \log(u/r)$ .

When  $p > 1$ , the standard approach is

$$Pr(r_P|a) = \frac{Pr(r_P, \hat{\zeta}_\tau|a)}{Pr(\hat{\zeta}_\tau|r_P, a)}$$

The result is again (4.3), but it includes a more general definition of  $r_P^*$

$$r_P^* = r_P + r_P^{-1} \log(C_\tau) + r_P^{-1} \log\left(\frac{\tilde{u}}{r_P}\right), \quad (4.4)$$

where

$$\begin{aligned} C_\tau &= \left| \frac{\partial^2 l(\hat{\theta}_\tau)}{\partial \zeta \partial \hat{\zeta}_\tau} \right| \left\{ |j_{\zeta\zeta}(\hat{\theta})| |j_{\zeta\zeta}(\tilde{\theta})| \right\}^{-1/2}, \\ \tilde{u}_\tau &= \left| \partial \{l_P(\hat{\theta}) - l_P(\theta)\} / \partial \hat{\tau} \right| \left| j_{\tau|\zeta}(\tilde{\theta}) \right|^{-1/2}, \end{aligned}$$

tilde indicates the estimation at  $\tilde{\theta} = (\tau, \hat{\zeta}_\tau)$  and  $j_{\zeta\zeta}(\theta)$  is the second order derivative  $-\partial^2 l(\theta)/\partial\zeta\partial\zeta^T$ . Instead  $j_{\tau|\zeta}(\theta)$  indicates the adjusted information for  $\tau$ , which is  $j_{\tau|\zeta}(\theta) = j_{\tau\tau}(\theta) - j_{\tau\zeta}(\theta)j_{\zeta\zeta}^{-1}(\theta)j_{\zeta\tau}(\theta)$ . It is common to express (4.4) as follows

$$r_P^* = r_P + NP + INF$$

where NP refers to *nuisance parameter* and INF to *information* adjustments.

In the Skovgaard approach, it is often unreasonable to compute the log-likelihood derivatives with respect to the data, but the best choice is to use a simulation of datasets under the model to approximate them. The approximations with second-order relative error are

$$\begin{aligned} \frac{\partial^2 l(\tau, \hat{\zeta}_\tau)}{\partial\theta\partial\hat{\theta}^T} &\cong \text{cov}_{\theta_0} \{U(\theta_1), U(\theta_0)^T\} i(\hat{\theta})^{-1} j(\hat{\theta}) \\ \frac{\partial \left\{ l(\hat{\tau}, \hat{\zeta}) - l(\tau, \hat{\zeta}_\tau) \right\}}{\partial\hat{\theta}} &\cong \text{cov}_{\theta_0} \{l(\theta_0) - l(\theta_1), U(\theta_0)\} i(\hat{\theta})^{-1} j(\hat{\theta}), \end{aligned}$$

where the parameters are considered as fixed values evaluated at  $\theta_0 = \hat{\theta}$  and  $\theta_1 = \tilde{\theta}$ . The covariances are computed without conditioning on an ancillary, and are approximated by Monte Carlo simulation from the fitted model.

Dealing with discrete data, when  $r^*$  is applied without any continuity correction, the  $p$ -value can be seen as an approximation of the mid- $p$ -value, i.e. the probability of strictly greater evidence plus one-half the probability of the observed result. It is shown that in relation to the discreteness it has more nearly a uniform distribution under the hypothesis than the continuity-corrected or exact  $p$ -value. More details can be found in the Section 6 of Pierce and Bellio (2017).

## 4.1 The likelihoodAsy Package

The *likelihoodAsy* package Bellio and Pierce (2016) is designed to simplify the application of the  $r^*$  statistic. It requires a user-supplied function which evaluates the likelihood, but since higher-order likelihood inference require model specification beyond the likelihood function, it is also required another user-supplied function which generates a sample under the assumed

parametric statistical model. The user may decide to supply also a function that evaluates the gradient of the log likelihood function, which may lead to safer computation and save computing time.

Both the functions for the log likelihood function and the one which generates a data set should have two arguments. The first argument, `theta`, is a numeric vector which contains the value of the parameter, whereas the other one, `data`, is a list that should contain all the data required for the model.

### 4.1.1 Examples

#### Astrology and Education

As already mentioned above, the package requires a function which returns the log-likelihood at a given parameter value and a function which simulates a dataset, and since to study contingency tables we use the Poisson loglinear model, they are given by

```
#Function which returns the Poisson log-likelihood function at a given
  parameter value
loglik.Pois <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X %*% theta)
  l <- sum(y * log(mu) - mu)
  return(l)
}

#Function which simulates a data set having Poisson distribution
gendat.Pois <- function(theta, data) {
  out <- data
  mu <- exp(data$X %*% theta)
  out$y <- rpois(n=n, lam=mu)
  return(out)
}
```

while the function which returns the score function is given by

---

```
#Function which returns the Poisson score function at a given parameter
  value
score.Pois <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X %**% theta)
  s <- y %**% data$X - t(mu) %**% data$X
  return(s)
}
```

The  $r^*$  statistic is obtained through the function `rstar` of the library `likelihoodAsy`.

In this case, using the Linear-by-Linear Association model to study the association between the variables, the parameter of interest is the one corresponding to the scores,  $\tau$ , and in particular we are interested in studying the independence hypothesis  $H_0 : \tau = 0$ . Regarding the scores, since the standardized ones were more computationally manageable, we chose to use those since this does not affect the inferential results.

```
library(likelihoodAsy)

rs <- rstar(data = data.astro, thetainit = theta.fix, floglik = loglik.
  Pois, fscore = score.Pois, fpsi = function(theta) theta[8], psival =
  0, datagen = gendat.Pois, trace = FALSE)
summary(rs)
```

Testing based on the  $r$  and  $r^*$  statistics

```
-----
Parameter of interest:      User-defined function
Skovgaard covariances computed with 1000 Monte Carlo draws
psi value under testing:
[1] 0
-----
```

Estimates

```
Maximum likelihood estimate of psi:
  uv
0.4823
Standard error of maximum likelihood estimate of psi:
```

```

[1] 0.05208
Maximum likelihood estimate of theta:
(Intercept)      astro11      astro12      degree11      degree12
      4.2135      -3.1178      -1.0682      1.2163      2.5173
      degree13      degree14      uv
      0.5945      0.9882      0.4823
Maximum likelihood estimate of theta under the null:
(Intercept)      astro11      astro12      degree11      degree12
      4.754e+00      -2.610e+00      -8.958e-01      1.755e-01      1.666e+00
      degree13      degree14      uv
      -1.169e-02      6.696e-01      -1.673e-21
-----
Test Statistics
Wald statistic P(r_wald<observed value; 1st order):
      uv      uv
      9.26      1.00
r statistic      P(r<observed value; 1st order):
      uv      uv
      10.04      1.00
r* statistic      P(r<observed value; 2nd order):
[1] 10.03      1.00
-----
Decomposition of high-order adjustment r*-r
NP adjustment      INF adjustment:
      uv
      -0.002722      -0.008041
-----

```

We see that in this case  $r$  and  $r^*$  lead to very similar results, since the sample size is very large (1793). Since the advantage in accuracy of  $r^*$  is more evident when the sample size is smaller and it is more reasonable to use samples having always the same size, for the simulation study we decided to generate the tables conditioning on the total count, and so using the multinomial distribution to generate them. However the Poisson model has been used to evaluate the log-likelihood, since, as shown in the previous Chapter, the inferential results are the same. This is true both for  $r$  and  $r^*$ .

```

#Function which returns the multinomial log-likelihood function at a given
  parameter value
loglik.Multinom <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  l <- dmultinom(y,prob=pi,log=TRUE)
  return(l)
}

#Function which generates a dataset having multinomial distribution
gendat.Multinom <- function(theta, data) {
  out <- data
  X <- data$X
  mu <- exp(X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  out$y <- as.vector(rmultinom(n=1, size=sum(data$y), prob=pi))
  return(out)
}

data.astro1 <- data.astro
data.astro1$X <- data.astro1$X[,-1]

rs1 <- rstar(data = data.astro1, thetainit = theta.fix[-1], floglik =
  loglik.Multinom, fpsci = function(theta) theta[7], psival = 0, datagen
  = gendat.Multinom, trace = FALSE)
summary(rs1)

Testing based on the r and r* statistics
-----
Test Statistics
Wald statistic P(r_wald<observed value; 1st order):
  uv      uv
9.26    1.00

```

```

r statistic      P(r<observed value; 1st order):
  uv      uv
10.04    1.00
r* statistic     P(r<observed value; 2nd order):
[1]  10.03    1.00

```

-----

Decomposition of high-order adjustment  $r^*-r$

NP adjustment INF adjustment:

```

      uv
-0.002723  -0.008041

```

-----

the only difference is that the intercept is removed because of the reasons explained in Section 2.3.

The computation of confidence intervals based on the  $r^*$  statistic, and also the  $r$  statistic, can be done by using the `rstar.ci` function.

```

rs.ci <- rstar.ci(data = data.astro, thetainit = theta.fix, floglik =
  loglik.Pois, fscore = score.Pois, fpsf = function(theta) theta[8],
  datagen = gendat.Pois)

```

```
summary(rs.ci)
```

Confidence interval calculations based on likelihood asymptotics

-----

Parameter of interest: User-defined function

Calculations based on a grid of 18 points

Skovgaard covariances computed with 1000 Monte Carlo draws

-----

1st-order

90%	95%	99%
( 0.3979 , 0.5693 )	( 0.3820 , 0.5862 )	( 0.3512 , 0.6197 )

2nd-order

90%	95%	99%
( 0.3972 , 0.5685 )	( 0.3813 , 0.5854 )	( 0.3505 , 0.6189 )

-----

Decomposition of high-order adjustment

Nuisance parameter adjustment (NP)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.005558	-0.005332	-0.005061	-0.005047	-0.004769	-0.004491
Information adjustment (INF)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.009356	-0.009259	-0.009141	-0.009133	-0.009012	-0.008886

Again, using the multinomial model the results are exactly the same.

```
rs.ci1 <- rstar.ci(data = data.astro1, thetainit = theta.fix[-1], floglik
  = loglik.Multinom, fpsci = function(theta) theta[7], datagen = gendat.
  Multinom, trace = FALSE)
```

Using smaller sample sizes more differences between  $r$  and  $r^*$  are expected. For instance, generating a sample having a similar structure to the observed one fixing the total count to 50 using the function `gentabMult`, the results are different.

```
#Function which simulates a contingency table having multinomial
  distribution at a given parameter value
gentabMult <- function(theta, X, s) {
  mu <- exp(X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  y <- rmultinom(n=1, size=s, prob=pi)
  return(as.vector(y))
}

s <- 50
y <- gentabMult(theta.fix[-1], X[,-1], s=s)
data1 <- list(y=y, X=X)

rs <- rstar(data = data1, thetainit = theta.fix, floglik = loglik.Pois,
  fscore = score.Pois, fpsci = function(theta) theta[8], psival = 0,
  datagen = gendat.Pois, trace = FALSE)

summary(rs)
```

```

Testing based on the r and r* statistics
-----
Parameter of interest:      User-defined function
Skovgaard covariances computed with 1000 Monte Carlo draws
psi value under testing:
[1] 0
-----

Estimates
Maximum likelihood estimate of psi:
  uv
0.7392
Standard error of maximum likelihood estimate of psi:
[1] 0.4264
Maximum likelihood estimate of theta:
(Intercept)      astro11      astro12      degree11      degree12
      0.3279      -4.4205      -1.2200          1.2206          3.1818
degree13      degree14          uv
      1.1959          0.8669          0.7392
Maximum likelihood estimate of theta under the null:
(Intercept)      astro11      astro12      degree11      degree12
  1.253e+00  -3.555e+00  -9.163e-01  -5.108e-01  1.758e+00
degree13      degree14          uv
  1.823e-01  3.365e-01  1.245e-20
-----

Test Statistics
Wald statistic P(r_wald<observed value; 1st order):
  uv      uv
1.7336  0.9585
r statistic      P(r<observed value; 1st order):
  uv      uv
1.9671  0.9754
r* statistic      P(r<observed value; 2nd order):
[1]  1.8842  0.9702
-----

Decomposition of high-order adjustment r*-r

```

```

NP adjustment  INF adjustment:
      uv
-0.01866  -0.06426
-----

rs.ci <- rstar.ci(data = data1, thetainit = theta.fix, floglik = loglik.
  Pois, fscore = score.Pois, fpsi = function(theta) theta[8], datagen =
  gendat.Pois)

summary(rs.ci)
plot(rs.ci)

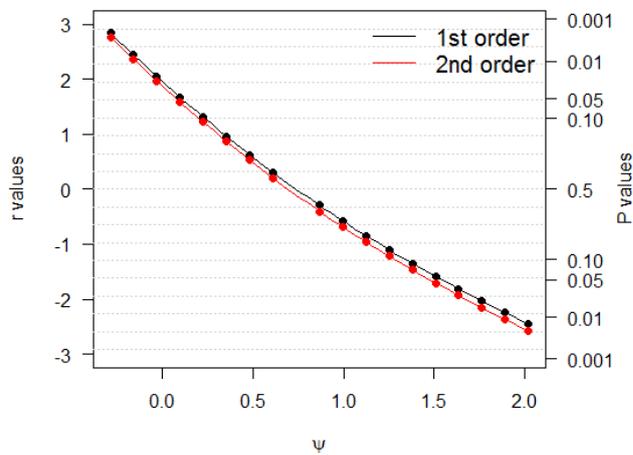
Confidence interval calculations based on likelihood asymptotics
-----
1st-order
      90%                95%                99%
( 0.1099 , 1.5391 )      ( 0.002422 , 1.718549 )      ( -0.1998 , 2.0961 )
2nd-order
      90%                95%                99%
( 0.08016 , 1.47380 )      ( -0.025 , 1.649 )      ( -0.2229 , 2.0178 )
-----

Decomposition of high-order adjustment
Nuisance parameter adjustment (NP)
      Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
-0.042330 -0.039360 -0.035100 -0.031300 -0.025050 -0.009604
Information adjustment (INF)
      Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
-0.08226  -0.07962  -0.07559  -0.07344  -0.06810  -0.05923
-----

plot(rs.ci)

```

We can see, for instance, that the 95% confidence interval obtained with  $r^*$  includes 0, whereas the one obtained with  $r$  does not contain it.



## Sexual Enjoyment Data

The output of `rstar` and `rstar.ci` using the data of Section 3.5.2, again testing for the null hypothesis  $H_0 : \tau = 0$ , is

```
rs <- rstar(data = data, thetainit = theta.fix, floglik = loglik.Pois,
  fscore = score.Pois, fpspsi = function(theta) theta[8], psival = 0,
  datagen = gendat.Pois, trace = FALSE)
```

```
summary(rs)
```

```
Testing based on the r and r* statistics
```

```
-----
Parameter of interest:      Independence test
Skovgaard covariances computed with 1000 Monte Carlo draws
psi value under testing:
[1] 0
-----
```

```
Estimates
```

```
Maximum likelihood estimate of psi:
```

```
uv
0.3847
```

```
Standard error of maximum likelihood estimate of psi:
```

```

[1] 0.126
Maximum likelihood estimate of theta:
(Intercept)      wife2      wife3      wife4      husband2
      0.92109      0.91844      0.43910      0.90351      0.07781
      husband3      husband4      uv
      -0.03949      0.35725      0.38467
Maximum likelihood estimate of theta under the null:
(Intercept)      wife2      wife3      wife4      husband2
      9.185e-01      8.473e-01      4.055e-01      1.012e+00      5.129e-02
      husband3      husband4      uv
      -2.246e-08      5.521e-01      -1.334e-21
-----
Test Statistics

Wald statistic P(r_wald<observed value; 1st order):
      uv      uv
3.0526  0.9989
r statistic      P(r<observed value; 1st order):
      uv      uv
3.2374  0.9994
r* statistic      P(r<observed value; 2nd order):
[1] 3.2008  0.9993
-----
Decomposition of high-order adjustment r*-r
NP adjustment      INF adjustment:
      uv
-0.01843  -0.01816
-----

rs.ci <- rstar.ci(data = data, thetainit = theta.fix, floglik = loglik.
  Pois, fscore = score.Pois, fpsci = function(theta) theta[8], datagen =
  gendat.Pois, trace = FALSE)

summary(rs.ci)

Confidence interval calculations based on likelihood asymptotics

```

```

-----
Parameter of interest:      User-defined function
Calculations based on a grid of 18 points
Skovgaard covariances computed with 1000 Monte Carlo draws
-----
1st-order
      90%                95%                99%
( 0.1845 , 0.6013 )      ( 0.1475 , 0.6452 )      ( 0.07593 , 0.73359 )
2nd-order
      90%                95%                99%
( 0.1788 , 0.5917 )      ( 0.1421 , 0.6352 )      ( 0.07117 , 0.72280 )
-----
Decomposition of high-order adjustment
Nuisance parameter adjustment (NP)
  Min.  1st Qu.  Median    Mean  3rd Qu.  Max.
-0.04632 -0.04202 -0.03587 -0.03493 -0.02833 -0.02064
Information adjustment (INF)

  Min.  1st Qu.  Median    Mean  3rd Qu.  Max.
-0.02804 -0.02638 -0.02414 -0.02388 -0.02152 -0.01891
-----

```

Here the results are slightly different but again quite close, and for the same reasons as before the choice was to try a sample with total count equal to 45, about half of the original sample size. Also for this example it is shown that the results using the multinomial are the same. Generating the table from the multinomial distribution, the results are

```

summary(rs)

Testing based on the r and r* statistics
-----
Parameter of interest:      User-defined function
Skovgaard covariances computed with 1000 Monte Carlo draws
psi value under testing:
[1] 0
-----

```

```

Estimates
Maximum likelihood estimate of psi:
  uv
0.3035
Standard error of maximum likelihood estimate of psi:
[1] 0.1971
Maximum likelihood estimate of theta:
(Intercept)      wife2      wife3      wife4      husb2
-0.3252      0.7623      0.6161      0.6515      0.9701
  husb3      husb4      uv
  0.7626      1.0432      0.3035
Maximum likelihood estimate of theta under the null:
(Intercept)      wife2      wife3      wife4      husb2
-4.055e-01      7.732e-01      6.931e-01      8.473e-01      9.555e-01
  husb3      husb4      uv
  7.885e-01      1.163e+00      -1.052e-21
-----
Test Statistics
Wald statistic P(r_wald<observed value; 1st order):
  uv      uv
1.5398      0.9382
r statistic      P(r<observed value; 1st order):
  uv      uv
1.5873      0.9438
r* statistic      P(r<observed value; 2nd order):
[1] 1.5502      0.9395
-----
Decomposition of high-order adjustment r*-r
NP adjustment      INF adjustment:
  uv
-0.01801      -0.01915
-----

summary(rs.ci)

```

Confidence interval calculations based on likelihood asymptotics

-----  
1st-order

90%	95%	99%
( -0.01078 , 0.64513 )	( -0.06976 , 0.71547 )	( -0.1851 , 0.8586 )

2nd-order

90%	95%	99%
( -0.01747 , 0.62715 )	( -0.07554 , 0.69623 )	( -0.1892 , 0.8367 )

-----

Decomposition of high-order adjustment

Nuisance parameter adjustment (NP)

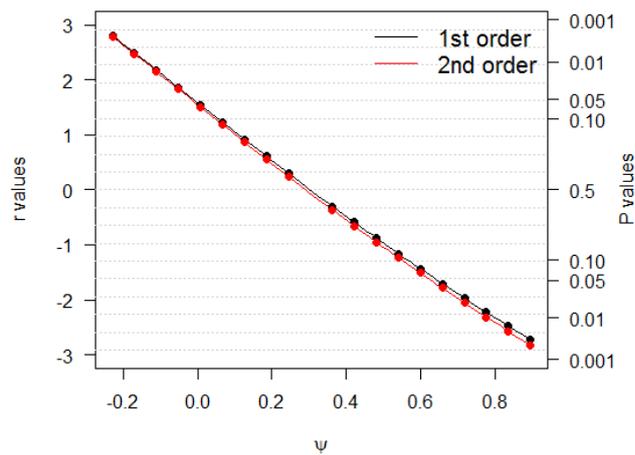
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.059100	-0.050420	-0.039000	-0.034960	-0.020270	-0.004426

Information adjustment (INF)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.03495	-0.03148	-0.02694	-0.02559	-0.02000	-0.01388

-----

`plot(rs.ci)`



The results using a reduced sample size are slightly different. However, there is not a big difference between the values of  $r$  and  $r^*$ .



# Chapter 5

## Simulation Study

In the simulation studies of this chapter, 10000 tables are generated from a multinomial distribution fixing the sample size of the contingency tables, whereas the evaluation of the log-likelihood and the calculation of  $r$  and  $r^*$  are done using the functions related to the Poisson distribution. As shown in Section 2.3 inference is not affected by this choice.

The datasets used as a basis for the simulation studies are the two datasets introduced in Sections 3.5.1 and 3.5.2, while the other simulations are done on simulated tables generated by a given vector of parameters. In particular the dimensions chosen for the simulated tables are  $5 \times 5$ ,  $6 \times 6$ ,  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$  and  $10 \times 10$ . For the dataset "Sexual Enjoyment Data"  $s$  is fixed equal to 45, about half of the original sample size, whereas for all the others simulations  $s$  is fixed equal to 50.

For each setting two simulations are performed: one in which the parameter of interest  $\tau$  is fixed to the value chosen for it, which in the case of the datasets is the maximum likelihood estimate while in the case of the tables is 0.3, and another one in which it is fixed to 0, i.e. under the independence model.

The results show the empirical coverage of  $r$  and  $r^*$ , the empirical coverage of the confidence intervals at a level of confidence of 90%, 95% and 99% and the average lengths of the confidence intervals. The graphics show the normal quantile plot for  $r$  and  $r^*$  and the distribution of the lengths of the confidence intervals.

Astrology and Education,  $\tau_0 = 0.48$ ,  $5 \times 3$ ,  $s = 50$

**Table 5.1:** Empirical coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.63%	2.21%	3.16%

**Table 5.2:** Empirical coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
95.23%	2.55%	2.22

**Table 5.3:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.18 (6.22, 4.6)	93.91 (3.24, 2.85)	98.81 (0.70, 0.49)
$r^*$	90.11 (4.84, 5.05)	94.6 (3.2, 2.20)	98.96 (0.5, 0.54)

**Table 5.4:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	1.141	1.368	1.824
$r^*$	1.114	1.336	1.781

Figure 5.1: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

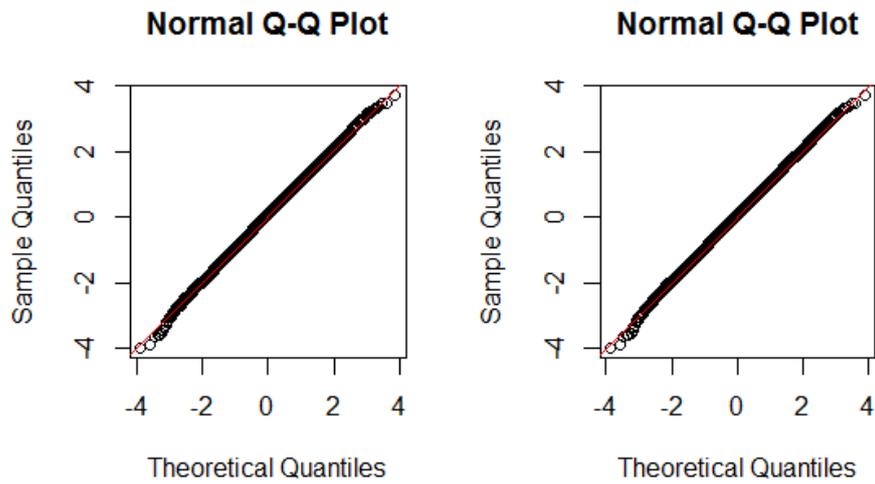
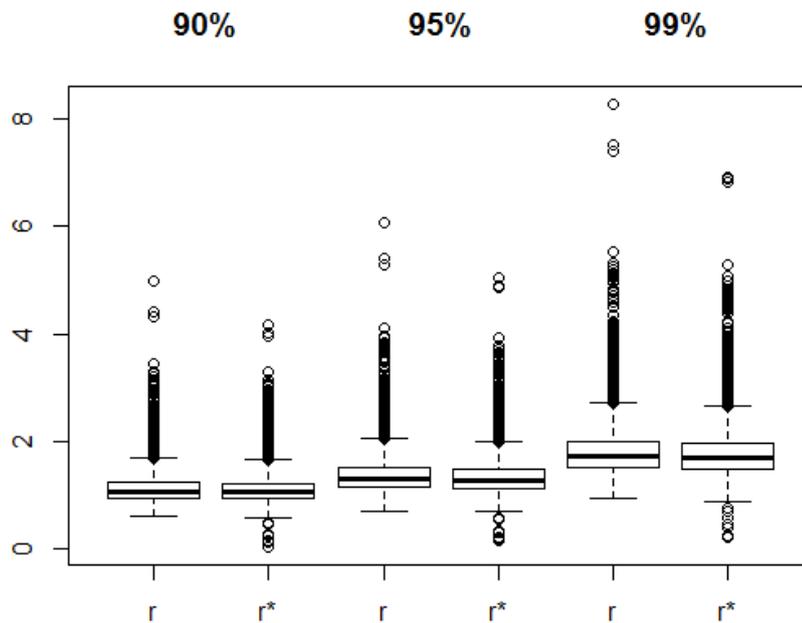


Figure 5.2: Distribution of the Lengths of the Confidence Intervals



Astrology and Education,  $\tau_0 = 0$ ,  $5 \times 3$ ,  $s = 50$

**Table 5.5:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.04%	2.87%	3.09%

**Table 5.6:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.93%	2.63%	2.44%

**Table 5.7:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.54 (5.93, 5.53)	93.91 (3.24, 2.85)	98.93 (0.54, 0.53)
$r^*$	89.43 (5.25, 5.32)	94.82 (2.61, 2.57)	98.99 (0.50, 0.51)

**Table 5.8:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	1.143	1.373	1.836
$r^*$	1.117	1.34	1.793

Figure 5.3: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

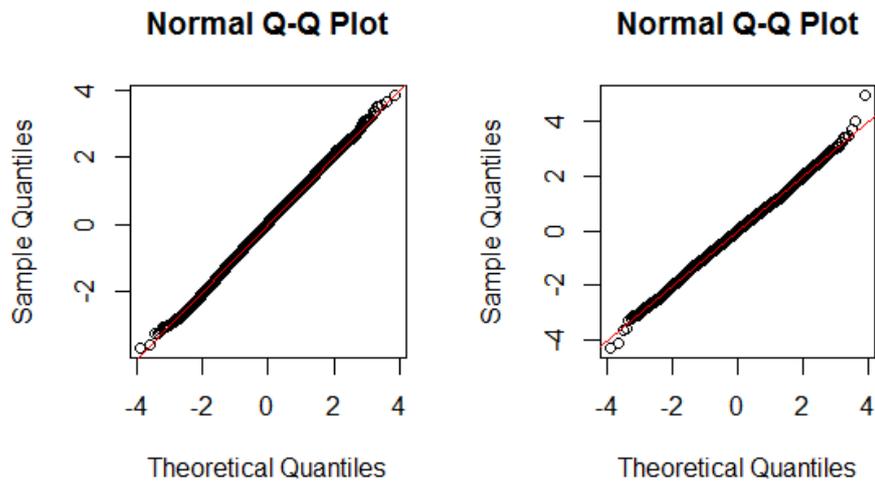
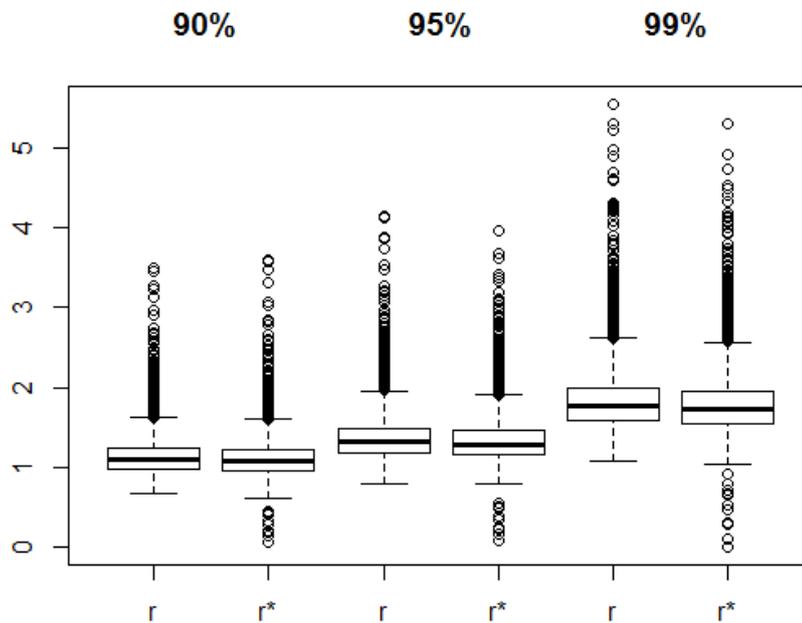


Figure 5.4: Distribution of the Lengths of the Confidence Intervals



Sexual Enjoyment Data,  $\tau_0 = 0.38$ ,  $4 \times 4$ ,  $s = 45$ **Table 5.9:** Empirical Coverage of  $r$ 

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
95.40%	1.65%	2.95

**Table 5.10:** Empirical Coverage of  $r^*$ 

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
96.36%	1.75%	1.89%

**Table 5.11:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	87.73 (7.29, 4.98)	95.41 (2.95, 1.64)	99.42 (0.42, 0.16)
$r^*$	91.06 (3.67, 5.27)	96.38 (1.88, 1.74)	99.51 (0.31, 0.18)

**Table 5.12:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.6467	0.7742	1.0287
$r^*$	0.6335	0.7584	1.0080

Figure 5.5: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

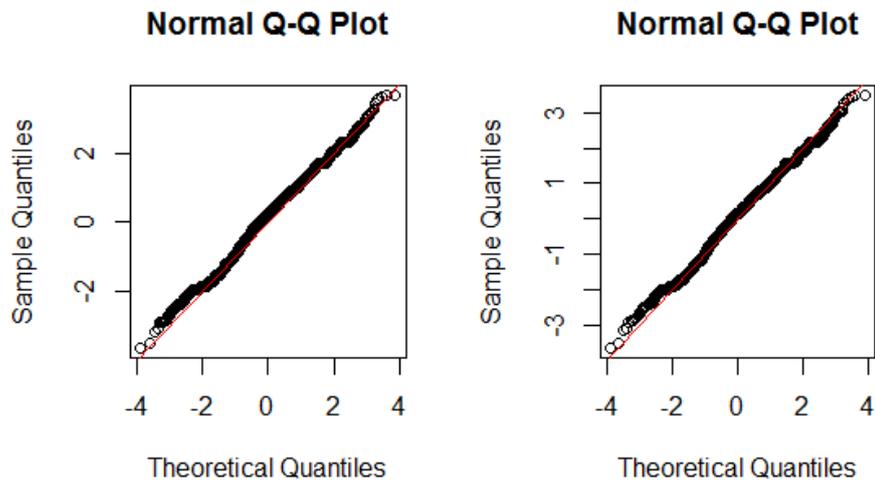
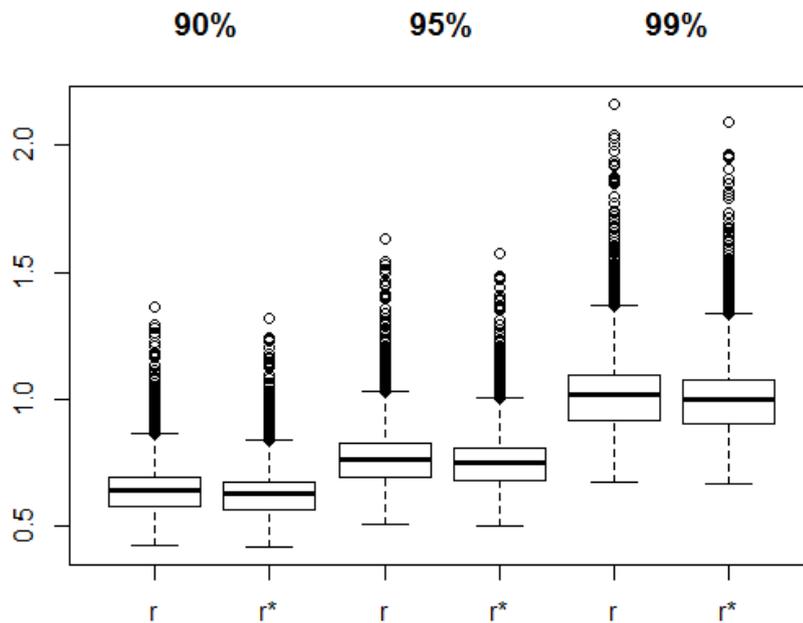


Figure 5.6: Distribution of the Lengths of the Confidence Intervals



Sexual Enjoyment Data,  $\tau_0 = 0$ ,  $4 \times 4$ ,  $s = 45$ **Table 5.13:** Empirical Coverage of  $r$ 

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.38%	2.75%	2.87%

**Table 5.14:** Empirical Coverage of  $r^*$ 

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.92%	2.5%	2.58%

**Table 5.15:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.25 (5.46, 5.29)	94.40 (2.85, 2.75)	98.75 (0.67, 0.58)
$r^*$	90.03 (5.08, 4.89)	94.95 (2.57, 2.48)	98.93 (0.59, 0.48)

**Table 5.16:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5785	0.6924	0.9198
$r^*$	0.5694	0.6815	0.9054

Figure 5.7: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

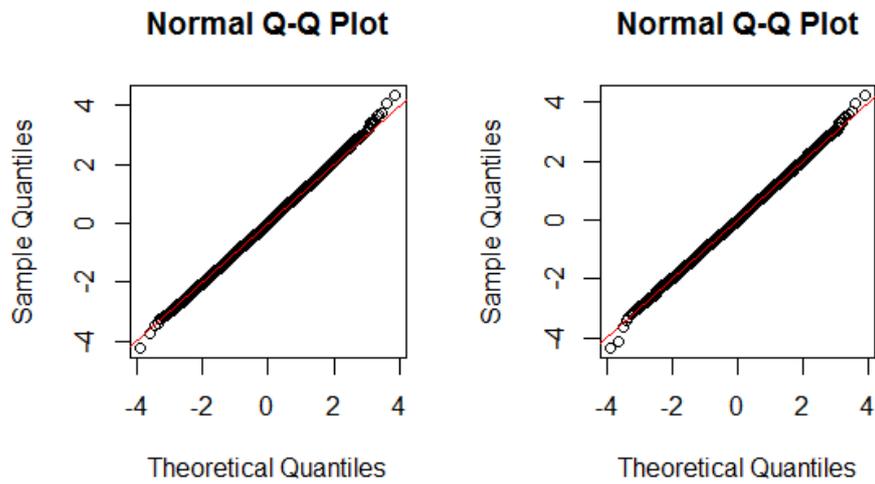
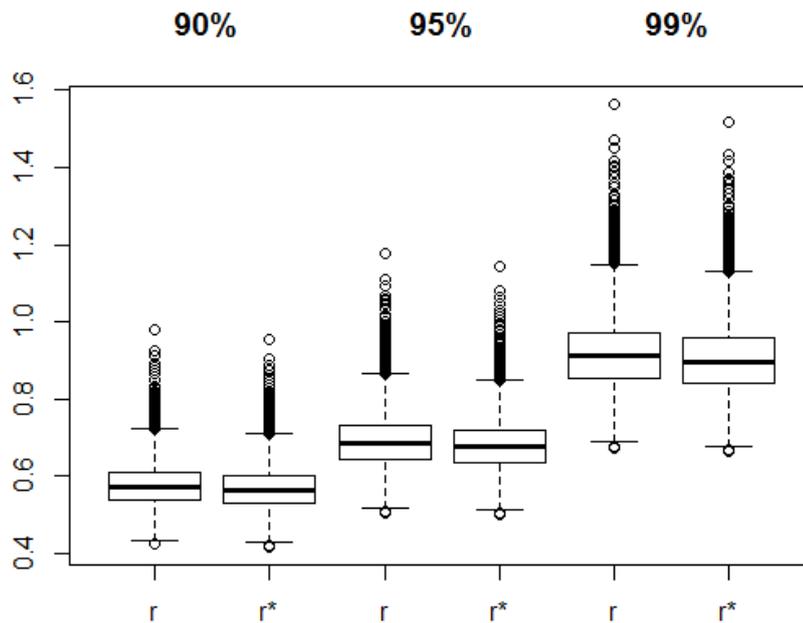


Figure 5.8: Distribution of the Lengths of the Confidence Intervals



$5 \times 5$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.17:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.09%	2.55%	3.36%

**Table 5.18:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.73%	2.63%	2.64%

**Table 5.19:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.77 (6.37, 4.86)	94.09 (3.35, 2.56)	98.75 (0.76, 0.49)
$r^*$	89.54 (5.33, 5.13)	94.73 (2.63, 2.64)	98.95 (0.55, 0.50)

**Table 5.20:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5985	0.7161	0.9505
$r^*$	0.5884	0.7041	0.9347

Figure 5.9: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

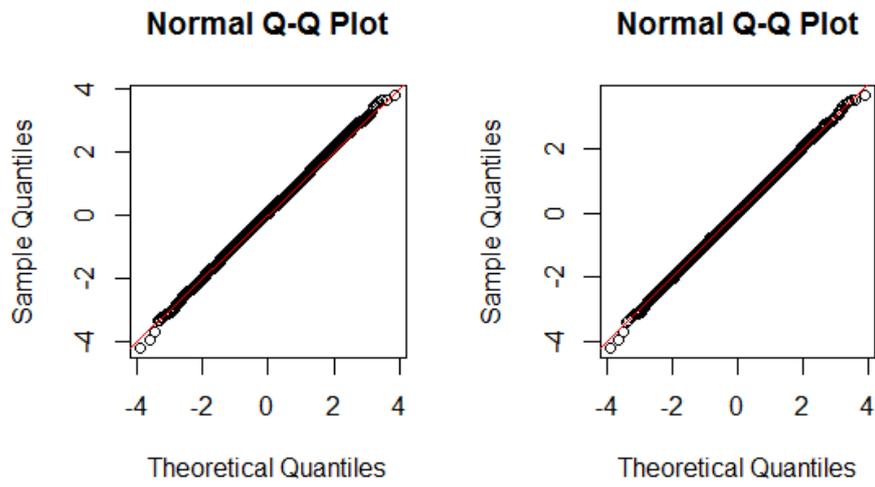
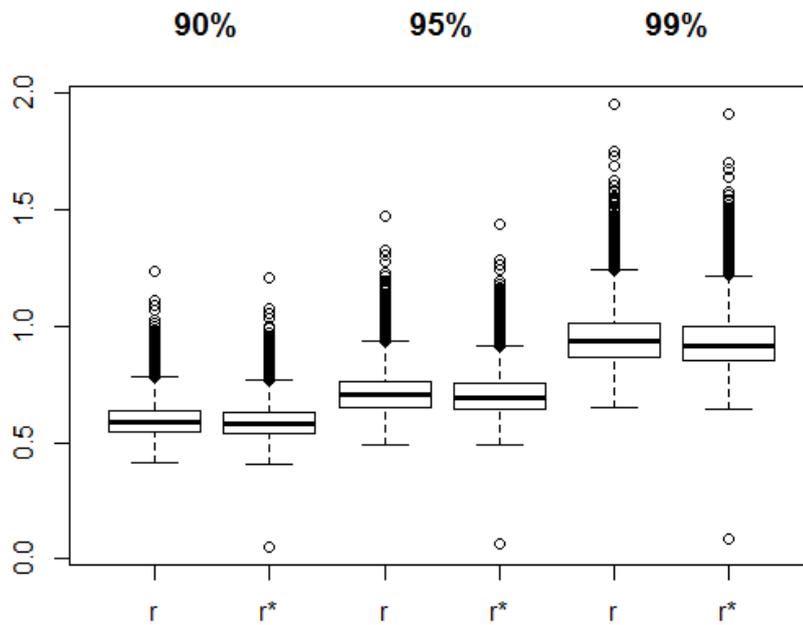


Figure 5.10: Distribution of the Lengths of the Confidence Intervals



$5 \times 5$  Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.21:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.35%	2.90%	2.75%

**Table 5.22:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.85%	2.65%	2.50%

**Table 5.23:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.25 (5.46, 5.29)	94.40 (2.86, 2.74)	98.75 (0.67, 0.58)
$r^*$	90.03 (5.08, 4.89)	94.95 (2.57, 2.48)	98.93 (0.59, 0.48)

**Table 5.24:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5644	0.6753	0.8966
$r^*$	0.5563	0.6657	0.8839

Figure 5.11: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

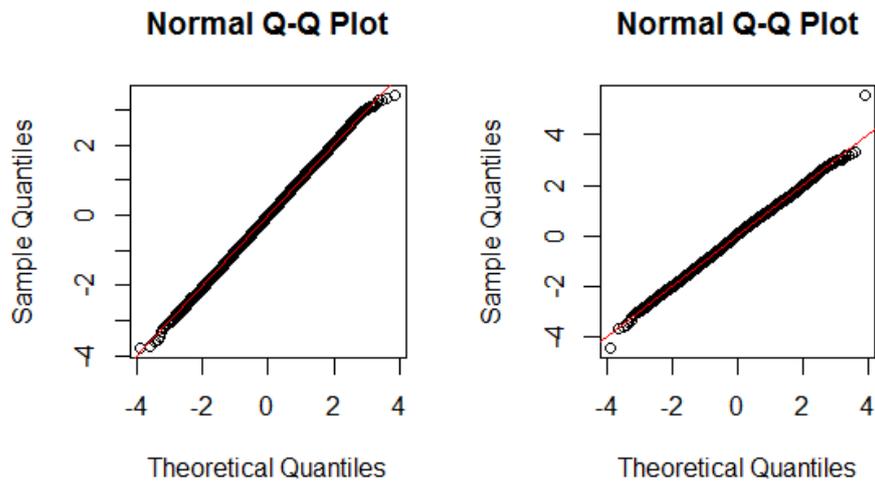
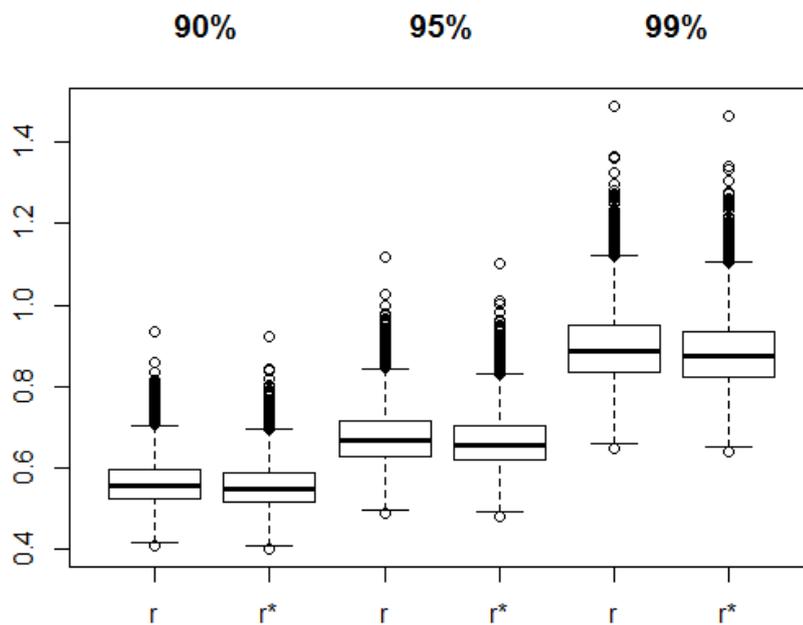


Figure 5.12: Distribution of the Lengths of the Confidence Intervals



$6 \times 6$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.25:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.26%	2.60%	3.14%

**Table 5.26:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.71%	2.70%	2.59%

**Table 5.27:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.04 (5.85, 5.11)	94.30 (3.11, 2.59)	98.77 (0.69, 0.54)
$r^*$	89.71 (4.76, 5.53)	94.78 (2.51, 2.71)	99.00 (0.44, 0.56)

**Table 5.28:** Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5784	0.6921	0.9188
$r^*$	0.5787	0.6805	0.9035

Figure 5.13: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

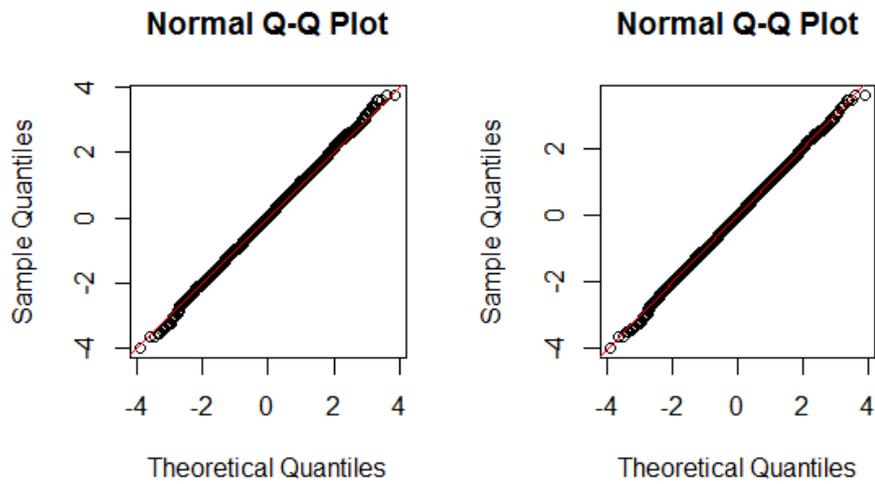
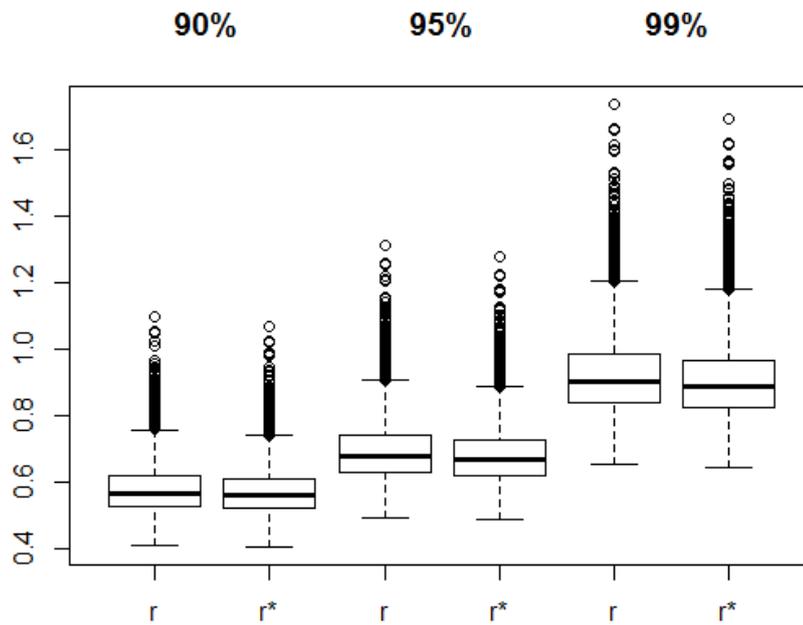


Figure 5.14: Distribution of the Lengths of the Confidence Intervals



$6 \times 6$  Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.29:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.37%	2.80%	2.83%

**Table 5.30:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.89%	2.55%	2.56%

**Table 5.31:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.97 (5.65, 5.38)	94.36 (2.83, 2.81)	98.88 (0.51, 0.61)
$r^*$	89.82 (5.20, 4.98)	94.91 (2.58, 2.51)	99.01 (0.45, 0.54)

**Table 5.32:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5432	0.6500	0.8630
$r^*$	0.5354	0.6407	0.8507

Figure 5.15: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

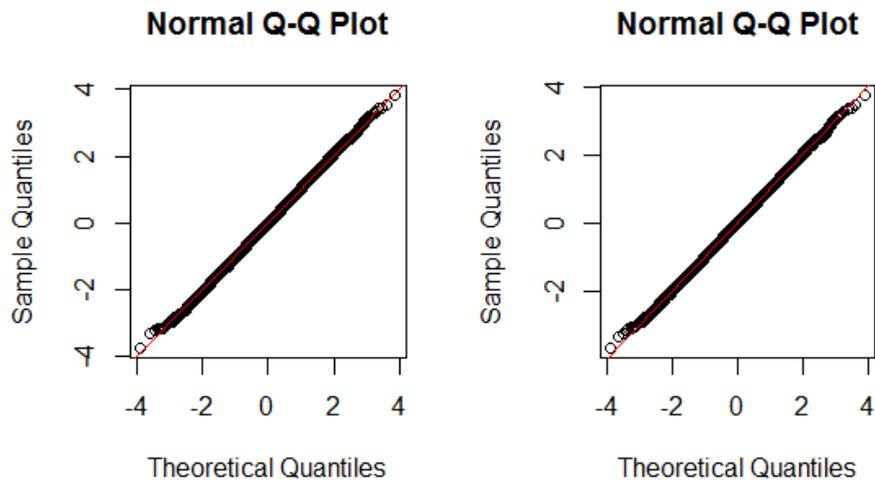
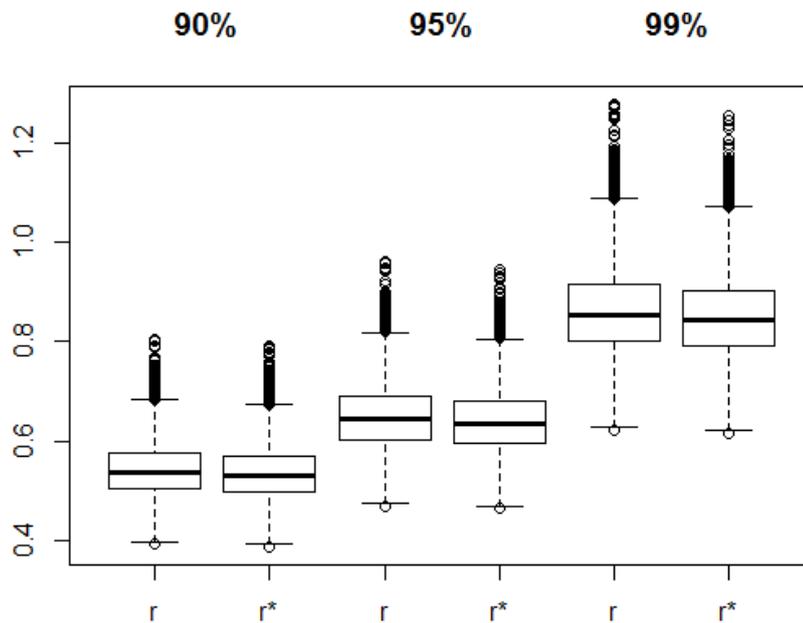


Figure 5.16: Distribution of the Lengths of the Confidence Intervals



$7 \times 7$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.33:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.68%	2.41%	2.91%

**Table 5.34:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
95.3%	2.48%	2.22%

**Table 5.35:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.53 (5.86, 4.61)	94.68 (2.91, 2.41)	98.72 (0.61, 0.67)
$r^*$	90.60 (4.58, 4.82)	95.29 (2.21, 2.5)	98.85 (0.47, 0.68)

**Table 5.36:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.6261	0.7492	0.9946
$r^*$	0.6155	0.7366	0.9781

Figure 5.17: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

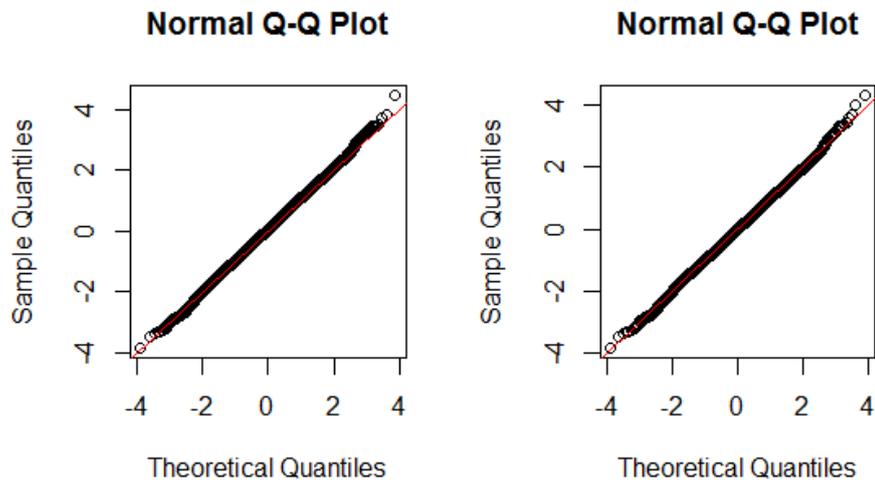
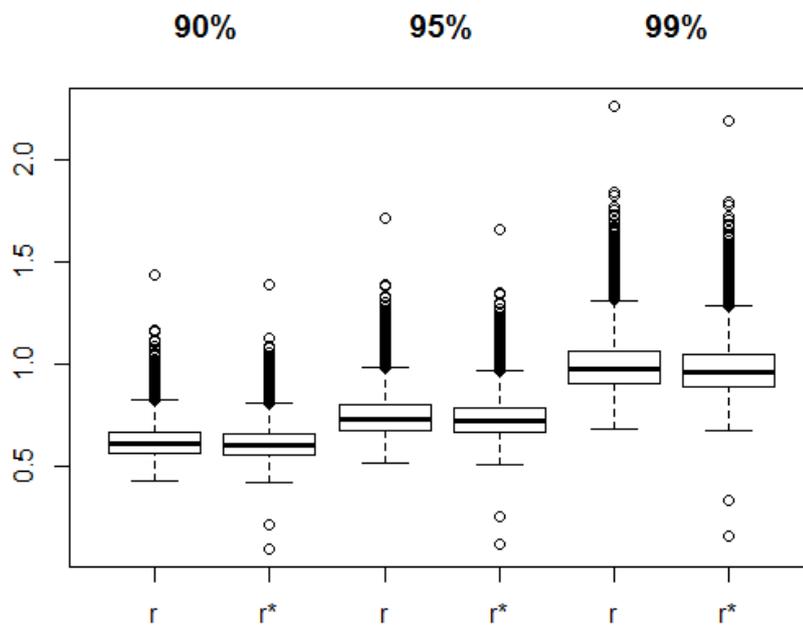


Figure 5.18: Distribution of the Lengths of the Confidence Intervals



$7 \times 7$  Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.37:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.49%	2.69%	2.82%

**Table 5.38:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
95.05%	2.45%	2.50%

**Table 5.39:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.12 (5.56, 5.32)	94.52 (2.82, 2.66)	98.89 (0.6, 0.51)
$r^*$	89.80 (5.19, 5.01)	95.09 (2.51, 2.4)	99.01 (0.50, 0.49)

**Table 5.40:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5949	0.7120	0.9457
$r^*$	0.5861	0.7015	0.9318

Figure 5.19: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

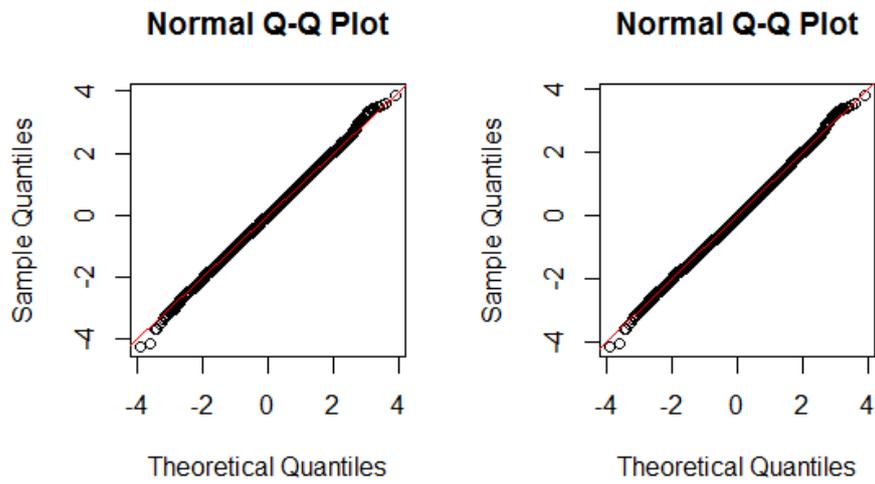
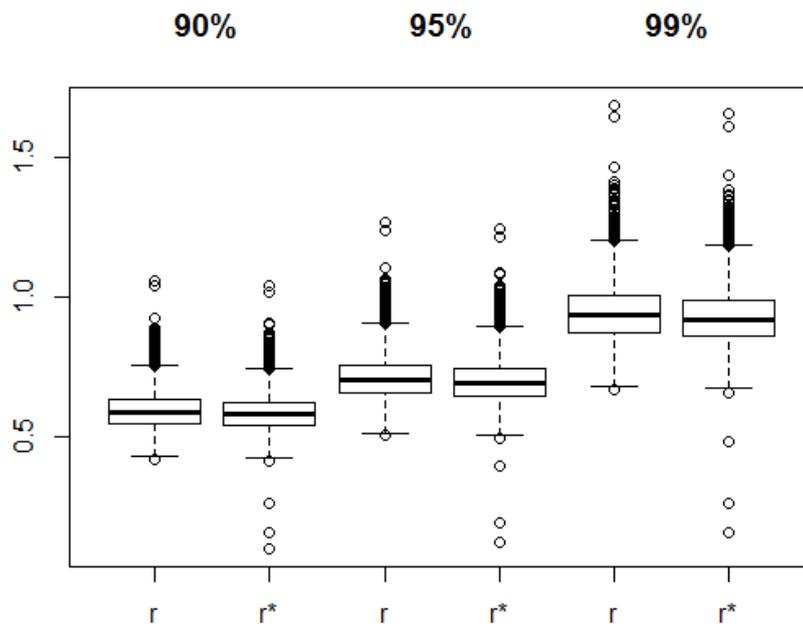


Figure 5.20: Distribution of the Lengths of the Confidence Intervals



$8 \times 8$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.41:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.52%	2.46%	3.02%

**Table 5.42:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
95.03%	2.58%	2.39%

**Table 5.43:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.33 (5.92, 4.75)	94.54 (3.04, 2.42)	98.82 (0.62, 0.56)
$r^*$	90.16 (4.78, 5.06)	95.01 (2.40, 2.59)	99.00 (0.42, 0.58)

**Table 5.44:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.6163	0.7374	0.9789
$r^*$	0.6058	0.7249	0.9625

Figure 5.21: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

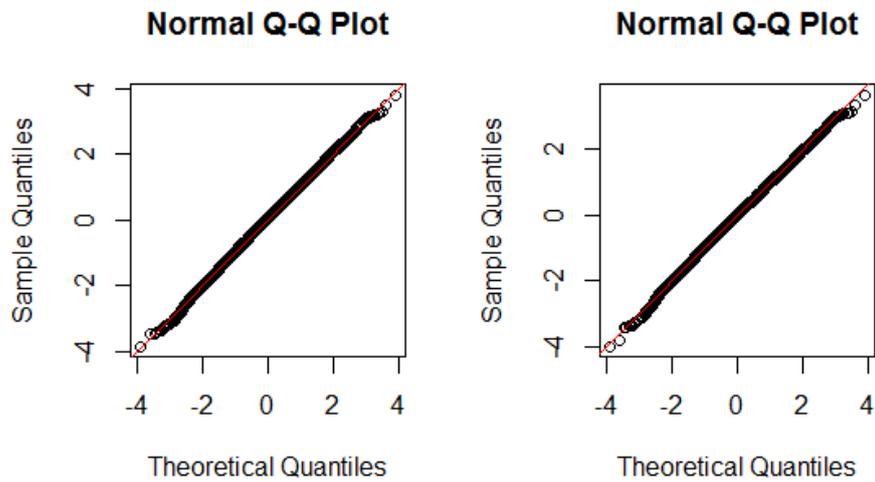
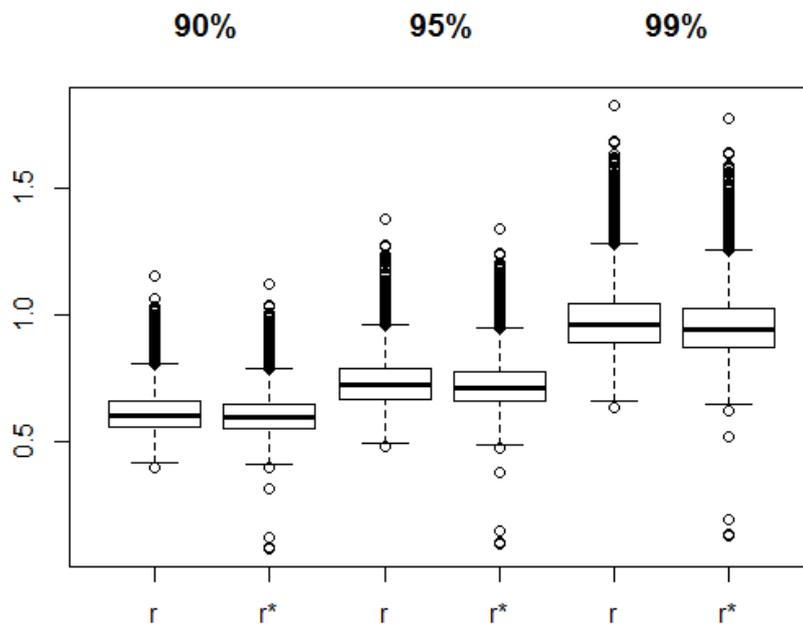


Figure 5.22: Distribution of the Lengths of the Confidence Intervals



$8 \times 8$  Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.45:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.07%	3.03%	2.9%

**Table 5.46:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.42%	2.77%	2.81%

**Table 5.47:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.82 (5.41, 5.77)	94.16 (2.85, 2.99)	98.71 (0.64, 0.65)
$r^*$	89.75 (4.98, 5.27)	94.73 (2.59, 2.68)	98.94 (0.55, 0.51)

**Table 5.48:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5868	0.7023	0.9327
$r^*$	0.5782	0.6921	0.9192

Figure 5.23: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

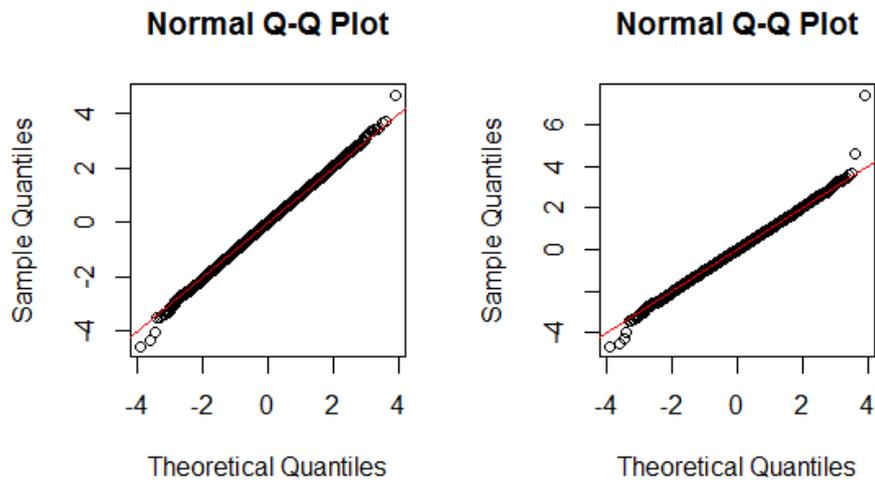
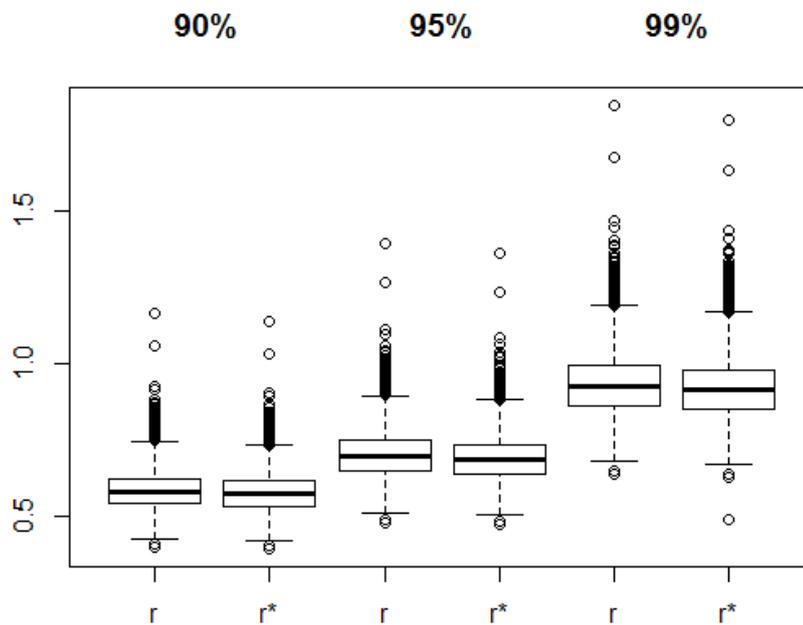


Figure 5.24: Distribution of the Lengths of the Confidence Intervals



$9 \times 9$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.49:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.69%	2.4%	2.91%

**Table 5.50:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.75%	2.73%	2.52%

**Table 5.51:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.53 (5.88, 4.59)	94.69 (2.99, 2.32)	98.83 (0.71, 0.46)
$r^*$	90.19 (4.81, 5.00)	95.28 (2.25, 2.47)	99.06 (0.45, 0.49)

**Table 5.52:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5661	0.6774	0.8991
$r^*$	0.5565	0.6659	0.8839

Figure 5.25: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

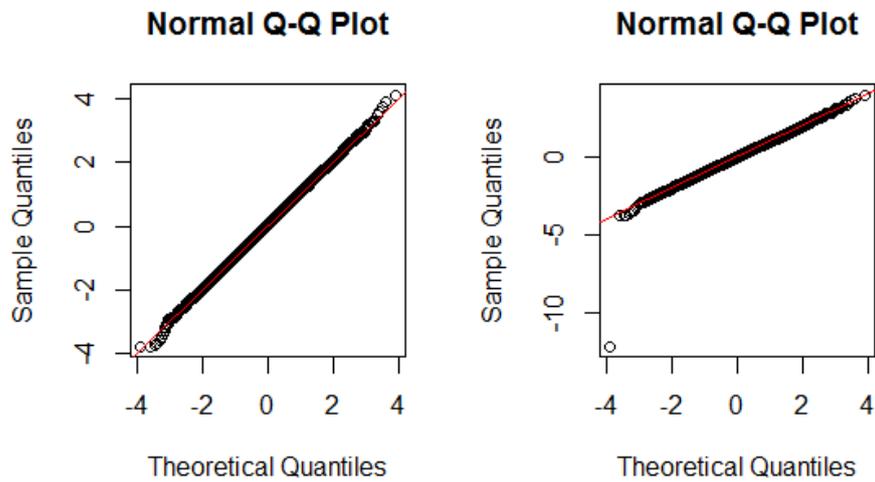
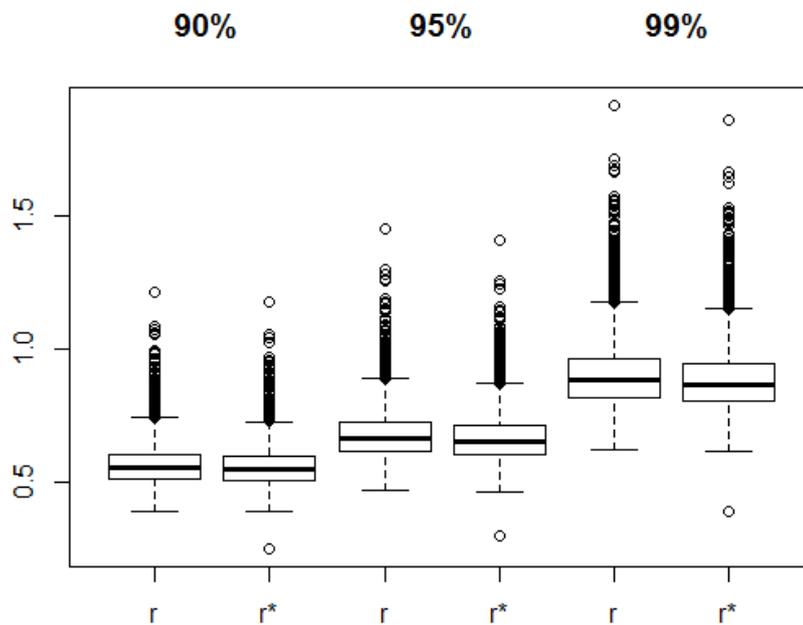


Figure 5.26: Distribution of the Lengths of the Confidence Intervals



$9 \times 9$  Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.53:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.48%	2.78%	2.74%

**Table 5.54:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.90%	2.63%	2.47%

**Table 5.55:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	89.14 (5.47, 5.39)	94.49 (2.71, 2.8)	98.84 (0.53, 0.63)
$r^*$	89.91 (5.02, 5.07)	94.92 (2.48, 2.6)	98.97 (0.49, 0.54)

**Table 5.56:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5321	0.6368	0.8455
$r^*$	0.5242	0.6273	0.8331

Figure 5.27: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

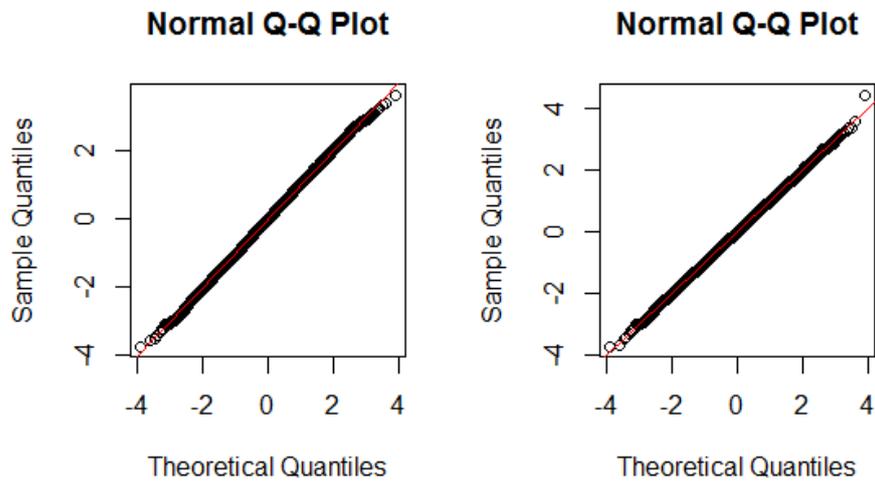
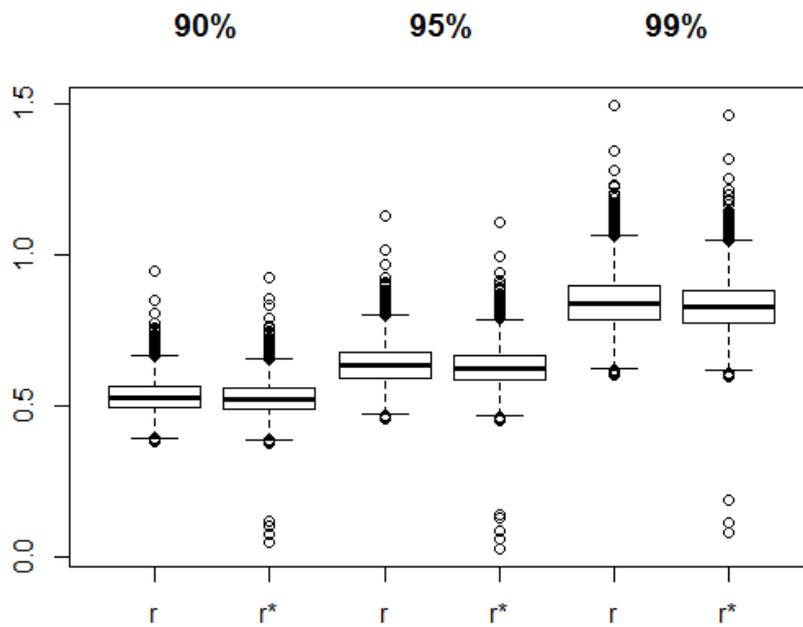


Figure 5.28: Distribution of the Lengths of the Confidence Intervals



$10 \times 10$  Table,  $\tau_0 = 0.3$ ,  $s = 50$

**Table 5.57:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.08%	2.55%	3.37%

**Table 5.58:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.73%	2.63%	2.64%

**Table 5.59:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.76 (6.37, 4.87)	94.09 (3.35, 2.56)	98.75 (0.76, 0.49)
$r^*$	89.54 (5.33, 5.13)	94.73 (2.63, 2.64)	98.95 (0.55, 0.50)

**Table 5.60:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5985	0.7161	0.9505
$r^*$	0.5883	0.7041	0.9347

Figure 5.29: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

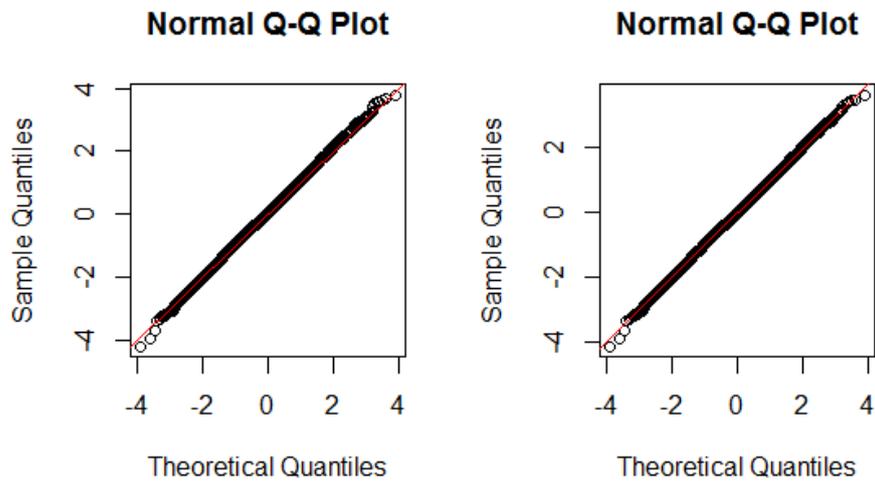
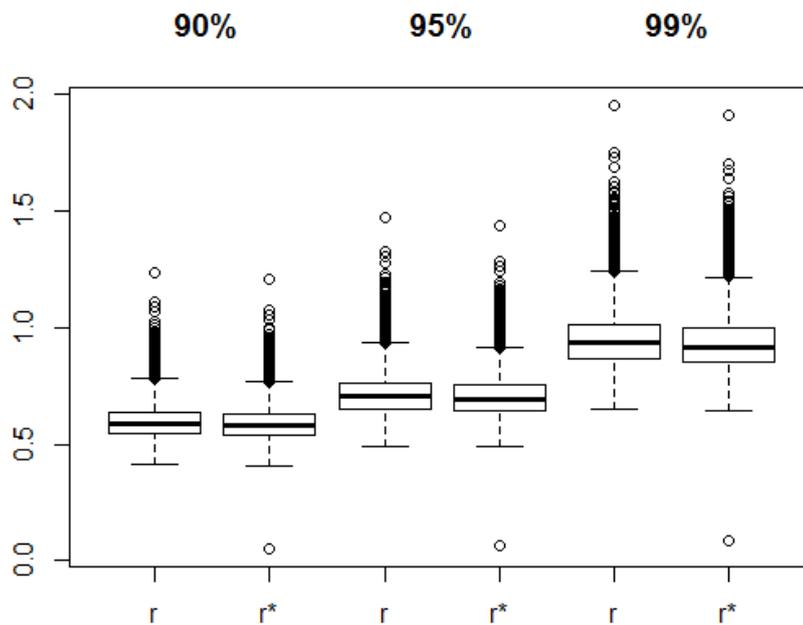


Figure 5.30: Distribution of the Lengths of the Confidence Intervals



10 × 10 Table,  $\tau_0 = 0$ ,  $s = 50$

**Table 5.61:** Empirical Coverage of  $r$

$ r  < z_{1-\alpha/2}$	$r < -z_{1-\alpha/2}$	$r > z_{1-\alpha/2}$
94.06 %	3.16%	2.78%

**Table 5.62:** Empirical Coverage of  $r^*$

$ r^*  < z_{1-\alpha/2}$	$r^* < -z_{1-\alpha/2}$	$r^* > z_{1-\alpha/2}$
94.33%	2.88%	2.79%

**Table 5.63:** Empirical Coverage of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	88.67 (5.39, 5.94)	93.99 (2.83, 3.18)	98.69 (0.58, 0.73)
$r^*$	89.38 (5.09, 5.53)	94.45 (2.71, 2.84)	98.93 (0.47, 0.60)

**Table 5.64:** Average Length of the Confidence Intervals

	CI 90%	CI 95%	CI 99%
$r$	0.5274	0.6311	0.8379
$r^*$	0.5195	0.6217	0.8254

Figure 5.31: Normal quantile plot of  $r$  (left panel) and  $r^*$  (right panel)

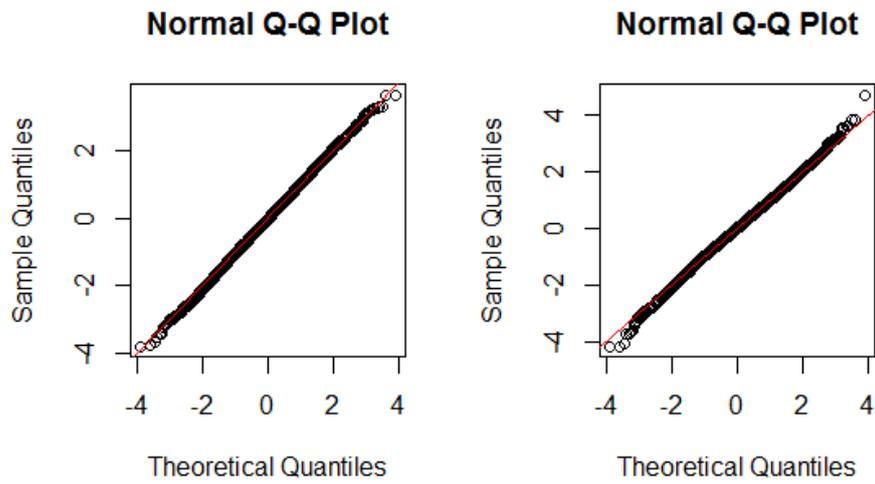
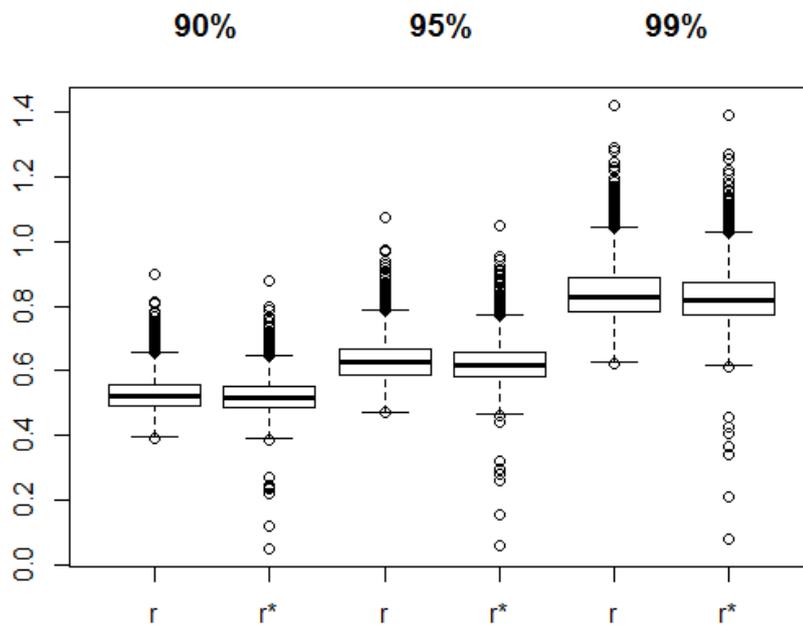


Figure 5.32: Distribution of the Lengths of the Confidence Intervals



## 5.1 Conclusions

The results show that  $r^*$  is slightly more accurate than  $r$  in all the empirical coverages and in particular, considering the empirical coverage of the confidence intervals, they show how  $r^*$  not only lead to a better coverage of the area between the quantiles, but also to a more symmetrical distribution of the quantities in the areas outside them. The different tables show that changes in the dimension of the table do not affect the accuracy of the results. This might be expected since the parameter of interest  $\tau$  is a canonical parameter of the Poisson distribution. Indeed, the profile likelihood coincides with the conditional likelihood for  $\tau$  and no nuisance parameter correction is needed. Therefore the only difference between  $r$  and  $r^*$  is given by the INF adjustment, which for this model is not very relevant.

From the lengths of the confidence intervals we see that the average length is always smaller for  $r^*$ , and that the distributions for  $r^*$  tend to be more symmetrical than the ones related to  $r$ .

# Bibliography

Agresti, A. (2010), *Analysis of Ordinal Categorical Data*. New Jersey: Wiley.

Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*. New Jersey: Wiley.

Azzalini, A. (2001), *Inferenza Statistica*. Milan: Springer Verlag.

Kolassa, J.E. and Tanner, M.A. (1994) Approximate Conditional Inference in Exponential Families via the Gibbs Sampler. *Journal of the American Statistical Association*, **89**, 697-702.

Pace, L. and Salvani, A. (2001), *Introduzione alla Statistica II*. Milan: Cedam.

Pierce, D.A. and Bellio, R. (2016), *likelihoodAsy: Functions for Likelihood Asymptotics*. R package version 0.45.

Pierce, D.A. and Bellio, R. (2017), Modern Likelihood-Frequentist Inference. *International Statistical Review*, to appear.



# Appendix A

## R Code

```
#Function which returns the log-likelihood function at a given parameter
  value
loglik.Pois <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X %*% theta)
  l <- sum(y * log(mu) - mu)
  return(l)
}

loglik.Multinom <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  l <- dmultinom(y,prob=pi,log=TRUE)
  return(l)
}

#Function which simulates a data set
gendat.Pois <- function(theta, data) {
  out <- data
```

```
mu <- exp(data$X %*% theta)
out$y <- rpois(n=n, lam=mu)
return(out)
}

gendat.Multinom <- function(theta, data) {
  out <- data
  X <- data$X
  mu <- exp(X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  out$y <- as.vector(rmultinom(n=1, size=sum(data$y), prob=pi))
  return(out)
}

#Function which simulates a contingency table at a given parameter value
gentabPois <- function(theta, X) {
  mu <- exp(X %*% theta)
  y <- rpois(n=n, lam=mu)
  return(y)
}

gentabella <- function(theta, X, s) {
  mu <- exp(X %*% theta)
  S <- sum(mu)
  pi <- mu/S
  y <- rmultinom(n=1, size=s, prob=pi)
  return(as.vector(y))
}

#Function which returns the score function at a given parameter value
score.Pois <- function(theta, data) {
  y <- data$y
  mu <- exp(data$X%*%theta)
  s <- y %*% data$X - t(mu) %*% data$X
  return(s)
}
```

```
}  
  
library(likelihoodAsy)  
  
### Example: Education and Belief About Astrology with Fit of Linear-by-  
Linear Association Model  
  
y <- c(98, 84, 23, 574, 286, 50, 122, 44, 4, 268, 57, 11, 148, 23, 1)  
#The number of cells of the table  
n <- length(y)  
#The sample size  
s <- 50  
  
#Astrology is Scientific: Not at all, Sort of, Very  
astro <- as.factor(rep(3:1,5))  
v <- as.numeric(astro)  
v <- scale(v)  
  
#Highest Degree: < High School, High School, Junior College, Bachelor,  
Graduate  
degree <- as.factor(c(rep(1,3), rep(2,3), rep(3,3), rep(4,3), rep(5,3)))  
u <- as.numeric(degree)  
u <- scale(u)  
  
uv <- u*v  
  
astro1 <- astro  
degree1 <- degree  
astro1 <- relevel(astro1,ref=3)  
degree1 <- relevel(degree1,ref=5)  
  
#Output for Fitting Linear-by-Linear Association Model  
LxL <- glm(y~astro1+degree1+uv, family=poisson)  
summary(LxL)  
  
y1 <- fitted(LxL)
```

```
theta.fix <- summary(LxL)$coefficients[,1]
theta.fix

X <- model.matrix(~astro1+degree1+uv)
X

data.astro <- list(y=y, X=X)

rs <- rstar(data = data.astro, thetainit = theta.fix, floglik = loglik.
  Pois, fscore = score.Pois, fpsf = function(theta) theta[8], psival =
  0, datagen = gendat.Pois, trace = FALSE)

summary(rs)

rs.ci <- rstar.ci(data = data.astro, thetainit = theta.fix, floglik =
  loglik.Pois, fscore = score.Pois, fpsf = function(theta) theta[8],
  datagen = gendat.Pois)

summary(rs.ci)
plot(rs.ci)

data.astro1 <- data.astro
data.astro1$X <- data.astro1$X[,-1]

rs1 <- rstar(data = data.astro1, thetainit = theta.fix[-1], floglik =
  loglik.Multinom, fpsf = function(theta) theta[7], psival = 0, datagen
  = gendat.Multinom, trace = FALSE)

summary(rs1)

rs.ci1 <- rstar.ci(data = data.astro1, thetainit = theta.fix[-1], floglik
  = loglik.Multinom, fpsf = function(theta) theta[7], datagen = gendat.
  Multinom, trace = FALSE)

summary(rs.ci1)
```

```
plot(rs.cil)

### Example: Sexual Enjoyment Data

y <- c(7, 2, 1, 2, 7, 8, 5, 8, 2, 3, 4, 9, 3, 7, 9, 14)
n <- length(y)
s <- 45

#Wife's response: Never or occasionally, Fairly Often, Very Often, Almost
  Always
wife <- as.factor(c(rep(1,4), rep(2,4), rep(3,4), rep(4,4)))
v <- as.numeric(wife)
v <- scale(v)

#Husband's response: Never or occasionally, Fairly Often, Very Often,
  Almost Always
husb <- as.factor(rep(1:4,4))
u <- as.numeric(husb)
u <- scale(u)

uv <- u*v
uv

X <- model.matrix(~wife+husb+uv)
X

#Output for Fitting Linear-by-Linear Association Model
LxL <- glm(y~wife+husb+uv, family=poisson)
summary(LxL)

y1 <- fitted(LxL)

theta.fix <- summary(LxL)$coefficients[,1]
theta.fix

data <- list(y=y, X=X)
```

```
rs <- rstar(data = data, thetainit = theta.fix, floglik = loglik.Pois,
  fscore = score.Pois, fpsi = function(theta) theta[8], psival = 0,
  datagen = gendat.Pois, trace = FALSE)

summary(rs)

rs.ci <- rstar.ci(data = data, thetainit = theta.fix, floglik = loglik.
  Pois, fscore = score.Pois, fpsi = function(theta) theta[8], datagen =
  gendat.Pois, trace = FALSE)

summary(rs.ci)
plot(rs.ci)

data1 <- data
data1$X <- data1$X[,-1]

rs1 <- rstar(data=data1, thetainit = theta.fix[-1], floglik = loglik.
  Multinom, fpsi = function(theta) theta[7], psival = 0, datagen =
  gendat.Multinom, trace=FALSE)

summary(rs1)

rs.ci1 <- rstar.ci(data = data1, thetainit = theta.fix[-1], floglik=loglik
  .Multinom, fpsi = function(theta) theta[7], datagen = gendat.Multinom,
  trace = FALSE)

summary(rs.ci1)
plot(rs.ci1)

#5x5 Table

n = 25
s = 50

x1 <- as.factor(rep(1:5,5))
```

```
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5)))
u1 <- as.numeric(x2)
u <- scale(u1)

uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.7, 0.6, 0.7, 0.8, 0.5, 0.4, 0.3, 0.5, 0.3)

#6x6 Table

n = 36
s = 50

x1 <- as.factor(rep(1:6,6))
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,6),rep(2,6),rep(3,6),rep(4,6),rep(5,6),rep(6,6)))
u1 <- as.numeric(x2)
u <- scale(u1)

uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.5, 0.6, 0.5, 0.6, 0.7, 0.6, 0.4, 0.5, 0.6, 0.8, 0.3)
```

```
#7x7 Table

n = 49
s = 50

x1 <- as.factor(rep(1:7,7))
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,7),rep(2,7),rep(3,7),rep(4,7),rep(5,7),rep(6,7),
  rep(7,7)))
u1 <- as.numeric(x2)
u <- scale(u1)

uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.5, 0.7, 0.5, 0.8, 0.5, 0.4, 0.3, 0.8, 0.5, 0.8, 0.5,
  0.6, 0.3)

#8x8 Table

n = 64
s = 50

x1 <- as.factor(rep(1:8,8))
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,8),rep(2,8),rep(3,8),rep(4,8),rep(5,8),rep(6,8),
  rep(7,8),rep(8,8)))
u1 <- as.numeric(x2)
u <- scale(u1)
```

```
uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.5, 0.7, 0.5, 0.8, 0.5, 0.4, 0.3, 0.8, 0.5, 0.8, 0.5,
              0.6, 0.5, 0.4, 0.3)

#9x9 Table

n = 81
s = 50

x1 <- as.factor(rep(1:9,9))
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,9),rep(2,9),rep(3,9),rep(4,9),rep(5,9),rep(6,9),
                 rep(7,9),rep(8,9),rep(9,9)))
u1 <- as.numeric(x2)
u <- scale(u1)

uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.5, 0.7, 0.5, 0.8, 0.5, 0.4, 0.3, 0.8, 0.5, 0.8, 0.5,
              0.6, 0.5, 0.4, 0.6, 0.7, 0.3)

#10x10 Table

n = 100
```

```
s = 50

x1 <- as.factor(rep(1:10,10))
v1 <- as.numeric(x1)
v <- scale(v1)

x2 <- as.factor(c(rep(1,10), rep(2,10), rep(3,10), rep(4,10), rep(5,10), rep
  (6,10), rep(7,10), rep(8,10), rep(9,10), rep(10,10)))
u1 <- as.numeric(x2)
u <- scale(u1)

uv <- u*v
uv

X <- model.matrix(~x1+x2+uv)
X

theta.fix <- c(1, 0.5, 0.7, 0.5, 0.8, 0.5, 0.4, 0.3, 0.8, 0.5, 0.8, 0.5,
  0.6, 0.5, 0.4, 0.6, 0.7, 0.5, 0.6, 0.3)

#Simulation

set.seed(3)

simulation <- function(N = 10000, s = 50, t = length(theta.fix)) {
  r <- numeric(N)
  rstar <- numeric(N)
  r.ci90 <- matrix(NA,ncol=2,nrow=N)
  r.ci95 <- matrix(NA,ncol=2,nrow=N)
  r.ci99 <- matrix(NA,ncol=2,nrow=N)
  rs.ci90 <- matrix(NA,ncol=2,nrow=N)
  rs.ci95 <- matrix(NA,ncol=2,nrow=N)
  rs.ci99 <- matrix(NA,ncol=2,nrow=N)

  for(i in 1:N) {
    y <- gentabMult(theta.fix[-1], X[, -1], s=s)
```

```
data1 <- list(y=y, X=X)
rs <- tryCatch({rstar(data = data1, thetainit = theta.fix, floglik =
  loglik.Pois, fscore = score.Pois, fpspsi = function(theta) theta[t],
  psival = theta.fix[t], datagen = gendat.Pois, trace = FALSE)}, error
  = function(e) "Error")

if(is.character(rs)) {
  r[i] <- NA
  rstar[i] <- NA
}
else {
  r[i] <- rs$r
  rstar[i] <- rs$rs
}

rs1 <- tryCatch({rstar.ci(data = data1, thetainit = theta.fix, floglik =
  loglik.Pois, fscore = score.Pois, fpspsi = function(theta) theta[t],
  datagen = gendat.Pois, trace = FALSE)}, error = function(e) "Error")

if(is.character(rs1)) {
  r.ci90[i,] <- rep(NA, 2)
  r.ci95[i,] <- rep(NA, 2)
  r.ci99[i,] <- rep(NA, 2)

  rs.ci90[i,] <- rep(NA, 2)
  rs.ci95[i,] <- rep(NA, 2)
  rs.ci99[i,] <- rep(NA, 2)
}
else {
  r.ci90[i,] <- rs1$CIr[1,]
  r.ci95[i,] <- rs1$CIr[2,]
  r.ci99[i,] <- rs1$CIr[3,]

  rs.ci90[i,] <- rs1$CIrs[1,]
  rs.ci95[i,] <- rs1$CIrs[2,]
```

```
    rs.ci99[i,] <- rs1$CIrs[3,]
  }

  print(i)
}
return(list(r, rstar, r.ci90, r.ci95, r.ci99, rs.ci90, rs.ci95, rs.ci99))
}

s <- simulation()

r <- unlist(s[1])
rstar <- unlist(s[2])
r.ci90 <- unlist(s[3])
r.ci95 <- unlist(s[4])
r.ci99 <- unlist(s[5])
rs.ci90 <- unlist(s[6])
rs.ci95 <- unlist(s[7])
rs.ci99 <- unlist(s[8])

r.ci90 <- matrix(r.ci90, 10000, 2)
r.ci95 <- matrix(r.ci95, 10000, 2)
r.ci99 <- matrix(r.ci99, 10000, 2)
rs.ci90 <- matrix(rs.ci90, 10000, 2)
rs.ci95 <- matrix(rs.ci95, 10000, 2)
rs.ci99 <- matrix(rs.ci99, 10000, 2)

alpha <- 0.05
z <- qnorm(1-alpha/2)

rstar <- rstar[!is.na(r)]
r <- r[!is.na(r)]

#r statistic

# -z(1-alpha/2) < r < z(1-alpha/2)
int <- (abs(r)<z)
```

```
# r < -z(1-alpha/2)
inf <- (r<(-z))
# r > z(1-alpha/2)
sup <- (r>z)

print(c(mean(int), mean(inf), mean(sup)))

#r* statistic

# -z(1-alpha/2) < r* < z(1-alpha/2)
ints <- (abs(rstar)<z)
# r* < -z(1-alpha/2)
infs <- (rstar<(-z))
# r* > z(1-alpha/2)
sups <- (rstar>z)

print(c(mean(ints, na.rm=TRUE), mean(infs, na.rm=TRUE), mean(sups,na.rm=
TRUE)))

plot(r,rstar)

qqnorm(r)
abline(0,1,col=2)
qqnorm(rstar)
abline(0,1,col=2)

threshold <- 0.1

r1 <- r[(abs(r)<threshold)|(is.na(rstar))]
rstar1 <- rstar[(abs(r)<threshold)|(is.na(rstar))]

r2 <- r[(abs(r)>=threshold)&(!is.na(rstar))]
rstar2 <- rstar[(abs(r)>=threshold)&(!is.na(rstar))]

qqnorm(rstar2)
plot(rstar2,r2)
```

```
#Predicted r*
#lm
fit <- lm(rstar2~r2)
summary(fit)
rstar3 <- summary(fit)$coefficients[,1]%*%t(cbind(rep(1,length(r1)),r1))
rstar3

#smooth.spline
library(apspline)

fit1 <- smooth.spline(r2, rstar2)
rstar4 <- predict(fit1, x=r1)$y

#"Corrected" r*
rstar[(abs(r)<threshold)|(is.na(rstar))] <- rstar4

par(mfrow=c(1,2))
qqnorm(r)
abline(0,1,col=2)
qqnorm(rstar)
abline(0,1,col=2)
par(mfrow=c(1,1))

infs <- 0
sups <- 0
ints <- 0

#Empirical Coverage of r and r*

#r statistic
print(c(mean(int), mean(inf), mean(sup)))

#r* statistic

# r* < -z(1-alpha/2)
```

```
infs <- (rstar<(-z))
# r* > z(1-alpha/2)
sups <- (rstar>z)
# -z(1-alpha/2) < r* < z(1-alpha/2)
ints <- (abs(rstar)<=z)

print(c(mean(ints), mean(infs), mean(sups)))

r.ci90 <- r.ci90[!is.na(r.ci90[,1]),]
r.ci95 <- r.ci95[!is.na(r.ci95[,1]),]
r.ci99 <- r.ci99[!is.na(r.ci99[,1]),]
rs.ci90 <- rs.ci90[!is.na(rs.ci90[,1]),]
rs.ci95 <- rs.ci95[!is.na(rs.ci95[,1]),]
rs.ci99 <- rs.ci99[!is.na(rs.ci99[,1]),]

#Length of the Confidence Intervals

#90
mean(r.ci90[,2]-r.ci90[,1])
mean(rs.ci90[,2]-rs.ci90[,1])

#95
mean(r.ci95[,2]-r.ci95[,1])
mean(rs.ci95[,2]-rs.ci95[,1])

#99
mean(r.ci99[,2]-r.ci99[,1])
mean(rs.ci99[,2]-rs.ci99[,1])

boxplot(r.ci90[,2]-r.ci90[,1], rs.ci90[,2]-rs.ci90[,1],
        r.ci95[,2]-r.ci95[,1], rs.ci95[,2]-rs.ci95[,1],
        r.ci99[,2]-r.ci99[,1], rs.ci99[,2]-rs.ci99[,1])

#Empirical Coverage of Confidence Intervals

t <- length(theta.fix)
```

```
#alpha <- 0.10
r90sx <- (r.ci90[,1]>theta.fix[t])
rs90sx <- (rs.ci90[,1]>theta.fix[t])

r90dx <- (r.ci90[,2]<theta.fix[t])
rs90dx <- (rs.ci90[,2]<theta.fix[t])

#r
print(c(1-mean(r90sx)-mean(r90dx),mean(r90sx),mean(r90dx)))
#rs
print(c(1-mean(rs90sx)-mean(rs90dx),mean(rs90sx),mean(rs90dx)))

#alpha <- 0.05
r95sx <- (r.ci95[,1]>theta.fix[t])
rs95sx <- (rs.ci95[,1]>theta.fix[t])

r95dx <- (r.ci95[,2]<theta.fix[t])
rs95dx <- (rs.ci95[,2]<theta.fix[t])

#r
print(c(1-mean(r95sx)-mean(r95dx),mean(r95sx),mean(r95dx)))
#rs
print(c(1-mean(rs95sx)-mean(rs95dx),mean(rs95sx),mean(rs95dx)))

#alpha <- 0.01
r99sx <- (r.ci99[,1]>theta.fix[t])
rs99sx <- (rs.ci99[,1]>theta.fix[t])

r99dx <- (r.ci99[,2]<theta.fix[t])
rs99dx <- (rs.ci99[,2]<theta.fix[t])

#r
print(c(1-mean(r99sx)-mean(r99dx),mean(r99sx),mean(r99dx)))
#rs
print(c(1-mean(rs99sx)-mean(rs99dx),mean(rs99sx),mean(rs99dx)))
```