Journal of Statistical Software
October 2014, Volume 61, Code Snippet 2. *http://www.jstatsoft.org/*

mdscore: An R Package to Compute Improved Score Tests in Generalized Linear Models

Antonio Hermes M. da Silva-Júnior
Universidade Federal do Rio G. do Norte

Damião Nóbrega da Silva
Universidade Federal do Rio G. do Norte

Silvia L. P. Ferrari
Universidade de São Paulo

Abstract

Improved score tests are modifications of the score test such that the null distribution of the modified test statistic is better approximated by the chi-squared distribution. The literature includes theoretical and empirical evidence favoring the improved test over its unmodified version. However, the developed methodology seems to have been overlooked by data analysts in practice, possibly because of the difficulties associated with the computation of the modified test. In this article, we describe the **mdscore** package to compute improved score tests in generalized linear models, given a fitted model by the `glm()` function in R. The package is suitable for applied statistics and simulation experiments. Examples based on real and simulated data are discussed.

Keywords: asymptotic test, Bartlett-type correction, chi-square distribution, Lagrange multiplier test, R, score test.

1. Introduction

The Lagrange multiplier or score test (Rao 1948) is a useful procedure in parametric statistical inference. The score test statistic is based on maximum likelihood estimation theory and one of its attractive features is the requirement of maximization of the likelihood function only under the null hypothesis. The null distribution of the test statistic is usually unknown or not obtained analytically. However, the traditional application of the score test employs the chi-squared distribution as an approximate reference distribution. Although this limiting approximation makes the test simple in practice, it may not yield a valid test procedure when the sample size is not sufficiently large.

One approach to improve the statistical properties of the score test when the sample size is not "large enough" was developed by Cordeiro and Ferrari (1991). These authors proposed a general formula to modify the score test statistic using a correction factor that is defined by a second-degree polynomial of the score statistic itself. The correction factor, known as a Bartlett-type correction, is obtained in such a way that the asymptotic error in the approximation between the null distribution of the modified statistic and the chi-squared distribution converges more rapidly to zero, as the sample size increases, than the corresponding error in the chi-squared approximation for the original score statistic. In the context of generalized linear models (Nelder and Wedderburn 1972; McCullagh and Nelder 1989) with known dispersion parameter, a Bartlett-type correction to the score statistic is given by Cordeiro, Ferrari, and Paula (1993). Analogous results for generalized linear models with unknown dispersion are given in Cribari-Neto and Ferrari (1995). An extension of these works for heteroskedastic generalized linear models, where the dispersion depends on covariates, is addressed by Cordeiro, Botter, Barroso, and Ferrari (2003). Another interesting extension is given in Ferrari and Cordeiro (1996), where the correction for the score test is derived for exponential family nonlinear models.

The aforementioned references demonstrate both theoretically and empirically better statistical properties for the modified score testing methodology over its original counterpart. Despite this superior performance, the former test has not yet been incorporated into the statistical practice. One possible explanation for this lack of attention might be the difficulties of computation of the terms involved in the factor that defines the modified test statistic. Such terms involve partial derivatives of the mean and variance functions. For a given generalized linear model, i.e., a model with specified linear predictor, variance and link functions, it can be tedious but possible to obtain a set of expressions for the computation of the correction factor using software that allows standard operations with vectors and matrices. However, in the analysis of a particular data set, it is sometimes necessary to try different link functions, or different random components for the data, or to modify the linear predictor. Naturally, each such change may require new expressions to compute the improved test statistic, which may be a quite undesirable approach in practice.

In this article, we describe the R (R Core Team 2014) package **mdscore** (da Silva-Junior and da Silva 2014) to produce improved score hypothesis tests for regression coefficients in generalized linear models. The computations follow the theory given by Cordeiro *et al.* (1993) and Cribari-Neto and Ferrari (1995), therefore allowing models with known and unknown dispersion parameter. This program is designed to be used in addition to the fit of the model by the `glm()` function in R. One of the distinguishing features of this program is that the users need not obtain the analytical expressions involved in the computations of the corrected score test. The required computations are performed using vector- and matrix-based operations, where the components of the required vectors and matrices are expressions involving partial derivatives of the mean and variance functions of the specified generalized linear model. This is properly handled by the package because of capabilities in R to work with vectors and matrices and to compute symbolic derivatives. A similar program is described in da Silva and Cordeiro (2009) for a correction with the same purpose as here for the likelihood ratio test.

This article is organized as follows: In Section 2, we present the theory involved in the computation of the corrected score test statistic. In Section 3, we provide some details about the **mdscore** package and its implementation. In Section 4, we present the results of a Monte Carlo simulation experiment in which the package was applied to investigate the properties of the two score statistics. Section 5 illustrates the use of the program with real data examples. Finally, in Section 6, we provide a summary for the article and concluding remarks.

2. Theoretical background

Let us consider the problem of hypothesis testing in generalized linear models. Let $y_1, \ldots, y_n$ be n independent random variables, with each y_l having probability density function or probability function

$$f(y; \theta_l, \phi) = \exp[\phi\{y\theta_l - b(\theta_l) + c(y_l)\} + a(y, \phi)], \tag{1}$$

where $b(\cdot)$, $c(\cdot)$ and $a(\cdot, \cdot)$ are known appropriate functions. The mean and the variance of y_l are $\mathsf{E}(y_l) = \mu_l = \mathrm{d}b(\theta_l)/\mathrm{d}\theta_l$ and $\mathsf{VAR}(y_l) = \phi^{-1}V_l$, respectively, where $V_l = \mathrm{d}\mu_l/\mathrm{d}\theta_l$ is called the variance function, and $\phi > 0$ is regarded as a precision parameter. The variance function characterizes the distribution. For the normal, gamma, inverse Gaussian, binomial and Poisson distributions we have (omitting index) $V = 1$, $V = \mu^2$, $V = \mu^3$, $V = \mu(1 - \mu)$ and $V = \mu$, respectively. The systematic component of the model is defined by $d(\mu_l) = \eta_l = \boldsymbol{x}_l^\top \boldsymbol{\beta}$ $(l = 1, \ldots, n)$, where $d(\cdot)$ is a known one-to-one twice differentiable link function, $\boldsymbol{x}_l^\top = (x_{l1}, \ldots, x_{lp})$ is a vector of covariates for the lth observation and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters to be estimated $(p < n)$. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ and assume that it is of full column rank, i.e., $\mathrm{rank}(\boldsymbol{X}) = p$. The null hypothesis being tested is $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$, which is to be tested against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$, where $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, with $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \ldots, \beta_p)^\top$.

The score statistic, say S_R, for testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ is given in a compact form in Equation (2.3) of Cordeiro *et al.* (1993). The limiting null distribution of S_R is χ_q^2 with an approximation error of order $\mathcal{O}(n^{-1})$. The approximation error can be reduced to $\mathcal{O}(n^{-3/2})$ if the corrected score statistic, S_R^*, proposed by Cordeiro and Ferrari (1991) is used in place of S_R:

$$S_R^* = S_R\{1 - (\gamma_3 + \gamma_2 S_R + \gamma_1 S_R^2)\}, \tag{2}$$

where γ_1, γ_2 and γ_3 are of order $\mathcal{O}(n^{-1})$. The multiplying factor is known as a Bartlett-type correction, and it is an extension of the well-known Bartlett correction for the likelihood ratio statistic. The coefficients γ_1, γ_2 and γ_3 in (2) are given by $\gamma_1 = A_3/\{12q(q + 2)(q + 4)\}$, $\gamma_2 = (A_2 - 2A_3)/\{12q(q + 2)\}$ and $\gamma_3 = (A_1 - A_2 + A_3)/(12q)$, where the constants A_1, A_2 and A_3 depend on the cumulants of the derivatives of the log-likelihood function up to the fourth order; see Appendix A.

3. Implementation

The improved score statistic (2) considers, as the score test, the estimation of the model parameters only under the null hypothesis. For a particular fitted model, the program applies partial symbolic differentiation and matrix-based operations to compute the coefficients of the Bartlett-type correction required by the improved statistic. These tasks are well handled by R as a result of its capabilities to obtain symbolic derivatives and perform vector or matrix operations through a wide range of functions.

The **mdscore** package contains a small set of functions. The main function is `mdscore()`, which yields the score and the corrected score tests of the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ against $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$ for model (1). The general syntax of `mdscore()` is as follows:

```
mdscore(fit.model, X1, phi = NULL)
```

where

- `fit.model`: an object that stores the results of a `glm()` fit of the model under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$, as described in Section 2;
- `X1`: the matrix with the columns of the model matrix $\boldsymbol{X}$ that correspond to the coefficients specified in the null hypothesis;
- `phi`: the precision parameter ϕ.

Some remarks about these arguments follow.

- The object `fit.model` is obtained using the usual options passed to the `glm()` function. For example, consider a generalized linear model for a response variable y with mean μ and linear predictor

  $$d(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5,$$

  where $x_1, x_2, \ldots, x_5$ are explanatory variables and β_0, $\beta_1, \ldots, \beta_5$ are regression coefficients. If we wish to test $\mathcal{H}_0 : \beta_4 = \beta_5 = 0$, then

  ```
  fit.model <- glm(y ~  x1 + x2 + x3, data = dataset,
    family = family-name("link"))
  ```

  where `dataset` is the object containing the data and `family-name("link")` is the R specification of the model random component and the link function d. If the test of interest is $\mathcal{H}_0 : \beta_1 = 2, \beta_3 = -3$, one could use

  ```
  fit.model <- glm(y ~  x2 + x4 + x5 + offset(2 * x1) + offset(-3 * x3),
    data = dataset, family = family-name("link"))
  ```

  where the purposes of the `offset()` function are to fix the terms `2 * x1` and `-3 * x3` with known coefficients equal to one and conform the estimation of the model parameters to the null hypothesis.

- One way to obtain the matrix `X1` is to first extract the model matrix from the `glm()` fit of the full model and then choose the columns that correspond to $\mathcal{H}_0$. For example, consider again the test of $\mathcal{H}_0 : \beta_4 = \beta_5 = 0$. The commands are:

  ```
  fit.full <- glm(y ~ x1 + x2 + x3 + x4 + x5, data = dataset,
    family = family-name("link"))
  X <- model.matrix(fit.full)
  X1 <- X[, c(5, 6)]
  ```

  Another method is to apply the `model.matrix()` function directly to the explanatory variables, i.e.,

  ```
  X1 <- model.matrix(~ x4 + x5 - 1, data = dataset)
  ```

The `-1` in the formula passed to `model.matrix()` prevents the addition of a column of ones into the matrix, which would be the case if $\beta_0 = 0$ was part of the null hypothesis:

- If the precision parameter is known, e.g., in a binomial, Poisson or exponential generalized linear model, it must be supplied in the place of `NULL` in the option `phi = NULL` of the `mdscore()` function. If ϕ is not informed, the value `NULL` is passed to the function and the theory is implemented with the maximum likelihood estimate of ϕ. When the `Gamma` family is specified in `fit.model`, the program uses the maximum likelihood estimate of ϕ provided by the R package **MASS** (Venables and Ripley 2002), which must be installed beforehand.
- When the family argument in the `glm()` object passed to `fit.model` is binomial, then the function `mdscore()` will produce the correct score and the improved score tests when the response variable is either 0 or 1. The situation in which the data are independent binomials based on $m > 1$ trials is not currently implemented.

The function `mdscore()` returns the following list of values:

- `Sr`: the value of the score statistic;
- `Srcor`: the value of the modified score statistic;
- `coef`: a vector with the coefficients A_1, A_2 and A_3;
- `n`: the total sample size;
- `df`: the number of degrees of freedom of the chi-squared approximations for the tests;
- `phi`: the precision parameter used in the computations.

Although one can easily extract `Sr` and `Srcor` from an object `obj` containing the results from a call to `mdscore()`, a cleaner summary is produced by the function `summary(obj)`, e.g., as

```
obj <- mdscore(fit.model, X1, phi = NULL)
summary(obj)
```

When the value of the score statistic is too large, the value of the improved test statistic may become negative. This is due to the quadratic form of the adjustment made. In such cases, the `mdscore()` function provides a warning message.

We illustrate the use of the functions in the package with some practical examples in Section 5.

4. Simulation experiment

In this section, we apply the **mdscore** package in a simulation experiment to compare the finite sample properties of the score (S_R) and modified score (S_R^*) test statistics under three different generalized linear models. Such an experiment is useful for checking that differences in the two tests are due to the adjustment made to the score statistics and not because of any violation of the model assumptions. To demonstrate the practical value of the two tests, we compare their properties with those of the Wald and likelihood ratio (LR) tests because they also use a chi-square reference distribution. The computation of the Wald and LR tests was made by using specific functions added to the package. Code to reproduce the experiment is provided in the supplementary material.

The experiment considered the three continuous models $y_l \sim \mathrm{N}(\mu_l, \phi^{-2})$, $y_l \sim \mathrm{Gamma}(shape = \phi, scale = \mu_l/\phi)$ and $y_l \sim \mathrm{InverseGaussian}(\mu_l, \phi)$, $l = 1, 2, \ldots, n$. We took the following systematic component for the models

$$d(\mu_l) = \eta_l = \beta_1 + \beta_2 x_{l2} + \beta_3 x_{l3} + \beta_4 x_{l4} + \beta_5 x_{l5} + \beta_6 x_{l6}, \tag{3}$$

where $d(\cdot)$ was given by the *identity* (normal model), *log* (gamma model) and *inverse* (inverse Gaussian model) transformations and the values of the explanatory variables were generated from the $\mathrm{Gamma}(shape = 2, scale = 0.2)$ distribution.

For each model of interest, we generated 30,000 independent samples of sizes $n = 20, 30$ and 40 for the response variable keeping the values of the explanatory variables fixed. The parameter ϕ was equal to 4 for the normal model and 40 for the other two models. For each sample size, the linear predictor was restricted to conform to the null hypothesis $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ and the non-null parameters $(\beta_1, \beta_5, \beta_6)$ were taken as $(20, 10, 10)$, $(2, 1, 1)$ and $(2/15, 1/15, 1/15)$ for the normal, gamma and inverse Gaussian models, respectively. The observations of the inverse Gaussian distribution were generated using the `rinv.gaussian()` function from the **VGAM** package (Yee and Wild 1996; Yee 2010). For each generated sample, the four test statistics S_R, S_R^*, Wald and LR were computed. The S_R and S_R^* statistics required the maximum likelihood estimates of the regression coefficients and the precision parameter ϕ under the model restricted by H_0. The Wald and LR statistics also required the maximum likelihood estimates of $\beta_1, \ldots, \beta_6$ and ϕ under the unconstrained model. The values of the four statistics were compared with the $1 - \alpha$ quantile of the chi-square distribution with three degrees of freedom, with nominal significance levels $\alpha = 0.10$, 0.05, and 0.01.

Table 1 gives the type I error rates (null rejection rates) of the four tests. For all three models and sample sizes, the results show that the Wald and LR tests have much higher rejection rates than the score test for all significance levels considered. For instance, under the normal model, the rates corresponding to the 5% nominal level are 18.8% (Wald), 12.9% (LR) and 6.5% (score), for $n = 20$, and 10.4% (Wald), 8.1% (LR) and 5.9% (score), for $n = 40$. Under the gamma model, the same rates are 19.0% (Wald), 13.1% (LR) and 6.6% (score), for $n = 20$, and 10.5% (Wald), 8.2% (LR) and 5.8% (score), for $n = 40$. The null rejection rates under the inverse Gaussian model are 21.2% (Wald), 12.3% (LR) and 6.9% (score), for $n = 20$, and 11.8% (Wald), 7.5% (LR) and 5.9% (score), for $n = 40$. The performances of these three tests improve as the sample increases, but the Wald and LR test seem to require a larger sample size than the score test to produce rejection rates near the nominal level.

The last column of Table 1 gives the rejection rates of the modified score test. It is seen that these rates are generally closer to the nominal levels than the rejection rates of the score test. Consider for example the case of $n = 20$ and nominal level 10%. The correction to the score statistic reduces the type I error rate from 14.2% to 10.2% in the normal model, from 14.3% to 10.3% in the gamma model and from 14.6% to 11.6% in the inverse Gaussian model. For the same sample size with the nominal level of 5%, the reductions in the type I error rates for the normal, gamma and inverse Gaussian models are from 6.5% to 4.8%, 6.6% to 5.0% and 6.9% to 5.5%, respectively. Smaller changes are observed in most cases when $\alpha = 0.01$.

Model	n	α (%)	Wald	LRT	Score	Mod. Score
Normal	20	10	26.9	21.0	14.2	10.2
		5	18.8	12.9	6.5	4.8
		1	8.9	4.0	0.6	0.6
	30	10	19.9	16.2	12.4	10.0
		5	12.6	9.3	6.0	5.1
		1	5.0	2.6	0.9	0.9
	40	10	17.0	14.5	11.8	10.1
		5	10.4	8.1	5.9	5.1
		1	3.6	2.1	0.9	0.9
Gamma	20	10	26.8	21.1	14.3	10.3
		5	19.0	13.1	6.6	5.0
		1	9.1	4.2	0.8	0.8
	30	10	20.4	16.6	12.8	10.2
		5	13.3	9.6	6.1	5.0
		1	4.9	2.6	0.9	0.9
	40	10	17.4	14.7	11.9	10.2
		5	10.5	8.2	5.8	5.0
		1	3.6	2.0	0.9	0.9
Inverse Gaussian	20	10	29.2	20.0	14.6	11.6
		5	21.2	12.3	6.9	5.5
		1	10.4	3.9	0.8	0.7
	30	10	22.2	15.4	12.4	10.9
		5	14.5	8.8	5.8	5.3
		1	5.9	2.2	0.8	0.8
	40	10	19.0	13.9	12.0	10.9
		5	11.8	7.5	5.9	5.4
		1	4.2	1.9	1.2	1.0

Table 1: Type I error rates of four statistics to test the hypothesis that $\beta_2 = \beta_3 = \beta_4 = 0$ in (3) based on 30,000 samples of size n from three generalized linear models.

5. Applications

We now provide two examples to illustrate the use of the **mdscore** package with real data. The first example is based on a gamma model fitted to a dataset with 34 observations. The second example illustrates the application of the inverse Gaussian distribution to a dataset with 30 observations. The results in each example can be reproduced with code available in the supplementary material. Before running this code, it is necessary to first load the **mdscore** package.

```
R> library("mdscore")
```

5.1. A gamma model with log-link

Ramsey and Schafer (2013) discuss a controlled experiment to test a disruption mechanism of the blood-brain barrier system. In this experiment, a set of 34 rats was initially inoculated with brain-tumor cancer cells. In periods varying from 9 to 11 days post inoculation, half of the animals were infused with a barrier disruption solution (BD) and the other half with a control solution (NS). After receiving a therapeutic antibody treatment, the rats were sacrificed and the amounts of antibody in the brain tumor and liver were determined. The data are available in the R package **Sleuth3** (Ramsey, Schafer, Sifneos, and Turlach 2013) under the name "case1102". The variables in this dataset are *Brain* (brain tumor count, per gm), *Liver* (liver count, per gm), *Time* (time of sacrifice after the therapeutic antibody treatment, in hours), *Treatment* (treatment received, BD or NS), *Days* (number of days between inoculation and the treatment administration), *Sex* (sex of the rat), *Weight* (rat initial weight, in grams), *Loss* (rat weight loss, in grams) and *Tumor* (tumor weight, in 10^{-4} grams). We can read this dataset by using the code

```
R> data("case1102", package = "Sleuth3")
R> d <- transform(case1102, TLrat = Brain / Liver, Ltime = log(Time),
+    Lwrat = log((Weight + Loss) / Weight),
+    Treat = factor(Treatment == "BD", labels = c("NS", "BD")))
```

where we added the ratio between the concentrations of antibody in the brain and in the liver (*TLrat*), the logarithm of the sacrifice time (*Ltime*), the logarithm of the ratio of the rat final weight to its initial weight (*Lwrat*) and a reformatted treatment variable.

Suppose Y_{ijk} denotes the value of a variable Y for the kth observation under the ith treatment, where i=0 ("NS"), 1 ("BD"), and the jth sacrifice time, j =1 (0.5 h), 2 (3 h), 3 (24 h) and 4 (72 h). Using the ratio of antibody concentrations as the response variable, one regression model that can applied to these data considers

$$TLrat_{ijk} \overset{\text{ind}}{\sim} \text{Gamma}(scale = \mu_{ij}/\phi, shape = \phi), \tag{4}$$

where

$$\begin{aligned}\log \mu_{ij} \;=\; & \beta_1 + \beta_2 Ltime_{ij} + \beta_3 I(Treat_{ij} = \text{``}BD\text{''}) + \beta_4 I(Treat_{ij} = \text{``}BD\text{''}) * Ltime_{ij} + \\ & \beta_5 Days_{ij} + \beta_6 I(Sex_{ij} == \text{``}Male\text{''}) + \beta_7 Lwrat_{ij} + \beta_8 Tumor_{ij} + \\ & \beta_9 I(Treat_{ij} = \text{``}BD\text{''}) * Lwrat_{ij}, \qquad i = 0, 1,\; j = 1, 2, 3, 4\end{aligned}$$

and $\phi > 0$. A summary of the fit of this model by maximum likelihood estimation is obtained by running

```
R> fitf <- glm(TLrat ~ Ltime * Treat + Days + Sex + Lwrat + Tumor +
+    Treat * Lwrat, data = d, family = Gamma("log"))
R> library("MASS")
R> summary(fitf, dispersion = gamma.dispersion(fitf))
R> gamma.shape(fitf)
```

The maximum likelihood estimates and standard errors of the nine regression coefficients and the estimate of the shape parameter of the model are given in Table 2. Figure 1 shows a normal quantile-quantile plot of the deviance residuals of the fitted model. This plot indicates that the gamma distribution is a reasonable random component for the model.

Suppose we are interested in the test of no interaction between the treatment and the logarithm of the ratio of the rat final weight to its initial weight, that is $H_0 : \beta_9 = 0$. The commands to obtain the score and modified score statistics are

Parameter	Estimate	SE
(Intercept)	−8.752	1.977
Ltime	1.076	0.072
TreatBD	1.203	0.267
Days	0.473	0.200
SexMale	0.114	0.249
Lwrat	7.359	8.621
Tumor	0.001	0.001
Ltime:TreatBD	−0.036	0.086
TreatBD:Lwrat	−24.707	10.012
Shape (ϕ)	4.584	1.074

Table 2: Maximum likelihood estimates for the parameters of model (4).

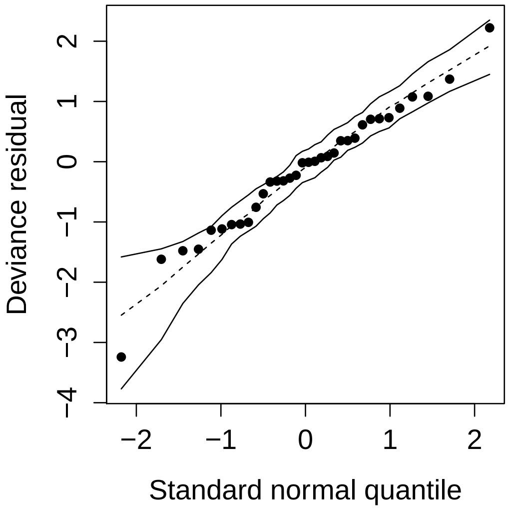

Figure 1: Normal quantile-quantile plot with an envelope for the deviance residuals of model (4).

```
R> X <- model.matrix(fitf)
R> fit0 <- glm(TLrat ~ Ltime * Treat + Sex + Lwrat + Tumor + Days,
+    data = d, family = Gamma("log"))
R> test <- mdscore(fit0, X1 = X[, 9], phi = NULL)
R> summary(test)
```

The results are shown in Table 3, which also presents the Wald and LR tests for the same hypothesis. These statistics and p values for the two tests can be obtained by

```
R> wald.test(fitf, terms = 9)
R> lr.test(fit1 = fit0, fit2 = fitf)
```

These results can be interpreted based on the findings in the simulation experiment in Section 4. The Wald test has the lowest p value, followed by the LR test that has the second lowest p value, because these tests tend to be more liberal than the score test. The score test

	Wald	LR	Score	Mod. Score
Statistic	6.0897	5.2293	4.3808	3.7226
p value	0.0136	0.0222	0.0363	0.0537

Table 3: Four statistical tests of the hypothesis $H_0 : \beta_9 = 0$ in model (4). The p values were computed using the chi-square distribution with one degree of freedom as the reference distribution.

still rejects the null hypothesis of no interaction at the 0.05 level, but after the correction the conclusion is reversed.

5.2. An inverse Gaussian model

As a second example, we take the data from a 5 × 2 factorial experiment given by Ostle and Mensing (1963). It is of interest to investigate how the impact strength of an insulating material is affected by the lot (I, II, III, IV, V) of the material and the type of specimen cut (lengthwise and crosswise). The experiment was replicated ten times, but for the purposes of illustrating the behavior of the test statistics under a smaller sample size, we randomly selected three replicates in each of the ten cells of this factorial experiment. The resulting dataset is stored in the object `strength` and the column `y` in this object refers to the impact strength observations.

An earlier analysis of the original dataset is given in Shuster and Miura (1972). In our analysis, we consider that the impact strength measurements of a given replicate corresponding to the ith cut and jth lot are independently distributed as inverse Gaussian distributions with means μ_{ij} and a fixed dispersion parameter. Suppose the linear predictor in the inverse link scale corresponds to the two-way interaction model

$$\mu_{ij}^{-1} = \tau_0 + \tau_i + \beta_j + (\tau\beta)_{ij}, \qquad i = 1, 2,\ j = 1, 2, \ldots, 5, \tag{5}$$

where $\tau_1 = 0$, $\beta_1 = 0$, and $(\tau\beta)_{11} = \cdots = (\tau\beta)_{15} = (\tau\beta)_{21} = 0$. This model can be fitted by the commands

```
R> fitf <- glm(y ~ cut * lot, data = strength,
+    family = inverse.gaussian("inverse"))
R> summary(fitf)
```

The adequacy of this model for the data is supported by the normal quantile-quantile plot of the deviance residuals given in Figure 2.

To test the hypothesis of no interaction between the two factors, that is $\mathcal{H}_0 : (\tau\beta)_{22} = (\tau\beta)_{23} = (\tau\beta)_{24} = (\tau\beta)_{25} = 0$, we first construct the model matrix

```
R> X <- model.matrix(fit.model, data = strength)
R> fit0 <- glm(y ~ cut + lot, data = strength,
+    family = inverse.gaussian("inverse"))
```

Inspecting the matrix `X`, it is observed that the interaction terms being tested in the null hypothesis are in columns 7 to 10. Therefore, the score and modified score tests are computed using

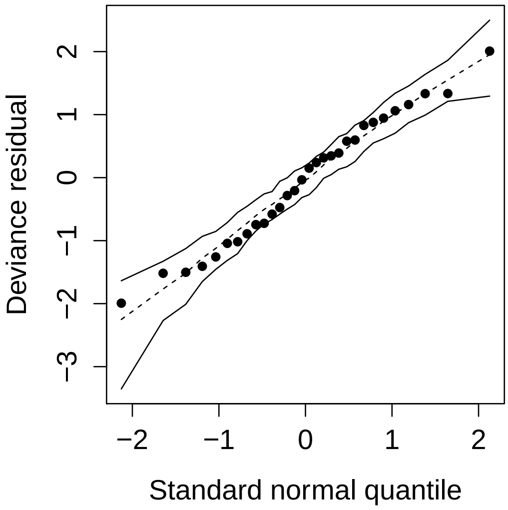

Figure 2: Normal quantile-quantile plot with envelope for the deviance residuals of model (5).

	Wald	LR	Score	Mod. Score
Statistic	10.5518	8.9572	7.6889	6.4179
p value	0.0321	0.0622	0.1037	0.1700

Table 4: Four statistical tests of the hypothesis $\mathcal{H}_0 : (\tau\beta)_{22} = (\tau\beta)_{23} = (\tau\beta)_{24} = (\tau\beta)_{25} = 0$ in model (5). The p values were computed using the chi-square distribution with four degrees of freedom as the reference distribution.

```
R> test <- mdscore(fit0, X1 = X[, 7:10])
R> summary(test)
```

The Wald and LR are easily computed with the functions

```
R> wald.test(fitf, terms = 7:10)
R> lr.test(fit1 = fit0, fit2 = fitf)
```

The results of the four tests are shown in Table 4. The Wald test suggests rejecting the null hypothesis at the 0.05 significance level, which is contrary to the other three tests. However, the ordering of the four statistics is consistent with those observed in the experiment in Section 4. Therefore, it seems sensible to conclude that the correction here not only helps support the conclusions of the LR and score tests but also presents much weaker evidence to reject H_0.

6. Concluding remarks

The R package **mdscore** computes improved score tests in the context of generalized linear models. The user does not need to develop expressions for the modified test statistic, as the program computes them automatically. The program basically requires the fit of the model under the null hypothesis by the `glm()` function and the model matrix associated with the parameters in the null hypothesis. One of the advantages of the package is the use of the R resources to handle symbolic differentiation to compute the derivatives that are necessary by adjustment theory. In addition to the modified score test, the program also computes the original score test and the corresponding p values, using the chi-square as the reference distribution. These summaries are useful in practice and can be used to compare the properties of the modified and original tests, such as type I error probabilities, in finite samples.

References

Cordeiro GM, Botter DA, Barroso LP, Ferrari SLP (2003). "Three Corrected Score Tests for Generalized Linear Models with Dispersion Covariates." *Statistica Neerlandica*, **57**(4), 391–409.

Cordeiro GM, Ferrari SLP (1991). "A Modified Score Test Statistic Having Chi-Squared Distribution to Order n^{-1}." *Biometrika*, **78**(3), 573–582.

Cordeiro GM, Ferrari SLP, Paula GA (1993). "Improved Score Tests for Generalized Linear Models." *Journal of the Royal Statistical Society B*, **55**(3), 661–674.

Cribari-Neto F, Ferrari SLP (1995). "Second Order Asymptotics for Score Tests in Generalised Linear Models." *Biometrika*, **82**(2), 426–432.

da Silva DN, Cordeiro GM (2009). "A Computer Program to Improve LR Tests for Generalized Linear Models." *Communications in Statistics – Simulation and Computation*, **38**(10), 2184–2197.

da Silva-Junior AHM, da Silva DN (2014). ***mdscore**: Improved Score Tests for Generalized Linear Models*. R package version 0.1-2, URL `http://www.R-project.org/package=mdscore`.

Ferrari SLP, Cordeiro GM (1996). "Corrected Score Tests for Exponential Family Nonlinear Models." *Statistics & Probability Letters*, **26**(1), 7–12.

Harris P (1985). "An Asymptotic Expansion for the Null Distribution of the Efficient Score Statistic." *Biometrika*, **72**(3), 653–659.

McCullagh P, Nelder JA (1989). *Generalized Linear Models*. Chapman & Hall/CRC, London.

Nelder JA, Wedderburn RWM (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society A*, **135**(3), 370–384.

Ostle B, Mensing RW (1963). *Statistics in Research: Basic Concepts and Techniques for Research Workers*. Iowa State University Press, Ames.

Ramsey FL, Schafer DW (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis*. 3rd edition. Brooks/Cole Cengage Learning, Australia.

Ramsey FL, Schafer DW, Sifneos J, Turlach BA (2013). ***Sleuth3**: Data Sets from Ramsey and Schafer's 'Statistical Sleuth (3rd ed)'*. R package version 0.1-6, URL `http://CRAN.R-project.org/package=Sleuth3`.

Rao CR (1948). “Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation.” *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**(1), 50–57.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Shuster JJ, Miura C (1972). “Two-Way Analysis of Reciprocals.” *Biometrika*, **59**(2), 478–481.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York. URL `http://www.stats.ox.ac.uk/pub/MASS4`.

Yee TW (2010). “The **VGAM** Package for Categorical Data Analysis.” *Journal of Statistical Software*, **32**(10), 1–34. URL `http://www.jstatsoft.org/v32/i10/`.

Yee TW, Wild CJ (1996). “Vector Generalized Additive Models.” *Journal of Royal Statistical Society B*, **58**(3), 481–493.

A. Technical material

Consider a parametric model indexed by an unknown k-vector parameter $\boldsymbol{\theta}$ and let $\ell(\boldsymbol{\theta})$ be the log-likelihood function. The parameter $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, the dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ being q and $k-q$, respectively. The interest lies in testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$, where $\boldsymbol{\theta}_1^0$ is a specified vector. Let $\boldsymbol{U}_{\boldsymbol{\theta}}$ and $\boldsymbol{K}_{\boldsymbol{\theta}}$ denote the score function and the Fisher information matrix for $\boldsymbol{\theta}$, respectively. The partition for $\boldsymbol{\theta}$ induces the corresponding partitions

$$\boldsymbol{U}_{\boldsymbol{\theta}} = (\boldsymbol{U}_1^\top, \boldsymbol{U}_2^\top)^\top, \quad \boldsymbol{K}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{pmatrix}, \quad \boldsymbol{K}_{\boldsymbol{\theta}}^{-1} = \begin{pmatrix} \boldsymbol{K}^{11} & \boldsymbol{K}^{12} \\ \boldsymbol{K}^{21} & \boldsymbol{K}^{22} \end{pmatrix},$$

where $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}$ is the inverse of $\boldsymbol{K}_{\boldsymbol{\theta}}$. Let $\widetilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^{0\top}, \widetilde{\boldsymbol{\theta}}_2^\top)^\top$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$ under $\mathcal{H}_0$. The Rao score statistic (S_R) for testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ is given by $S_R = \widetilde{\boldsymbol{U}}_1^\top \widetilde{\boldsymbol{K}}^{11} \widetilde{\boldsymbol{U}}_1$, where a tilde indicates evaluation at $\widetilde{\boldsymbol{\theta}}$. The limiting distribution of S_R is χ_q^2 under $\mathcal{H}_0$ with an approximation error of order $\mathcal{O}(n^{-1})$. The null hypothesis is rejected for a given nominal level, α, if the test statistic exceeds the upper $1-\alpha$ quantile of the χ_q^2 distribution. Clearly, S_R only involves estimation under $\mathcal{H}_0$.

The corrected score statistic proposed by Cordeiro and Ferrari (1991) is given by (2). It involves three coefficients, γ_1, γ_2 and γ_3, which depend on three other quantities, A_1, A_2 and A_3. General expressions for the A's involving cumulants of log-likelihood derivatives are given by Harris (1985); see also Equations (3.2)–(3.4) in Cordeiro *et al.* (1993). The coefficients A_1, A_2 and A_3 are functions of joint cumulants of log-likelihood derivatives up to the fourth order. Whenever they depend on unknown parameters, they can be evaluated at $\widetilde{\boldsymbol{\theta}}$. The null distribution of S_R^* is χ_q^2 with approximation error reduced from order $\mathcal{O}(n^{-1})$ to $\mathcal{O}(n^{-3/2})$.

To present the formulas for the coefficients A_1, A_2 and A_3 in generalized linear models, which enables us to obtain the corrected score statistic (2), it is useful to introduce some notation. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ be partitioned following the partition of the vector $\boldsymbol{\beta}$, $\boldsymbol{Z} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1}\boldsymbol{X}^\top$, $\boldsymbol{Z}_2 = \boldsymbol{X}_2(\boldsymbol{X}_2^\top \boldsymbol{W} \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top$, $\boldsymbol{W} = \operatorname{diag}(w_1, \ldots, w_n)$, $\boldsymbol{F} = \operatorname{diag}(f_1, \ldots, f_n)$, $\boldsymbol{G} = \operatorname{diag}(g_1, \ldots, g_n)$, $\boldsymbol{B} = \operatorname{diag}(b_1, \ldots, b_n)$ and $\boldsymbol{H} = \operatorname{diag}(h_1, \ldots, h_n)$, with

$$\begin{aligned}
w_l = w(\mu_l) &= \frac{1}{V_l}\left(\frac{d\mu_l}{d\eta_l}\right)^2, \\
f_l = f(\mu_l) &= \frac{1}{V_l}\frac{d\mu_l}{d\eta_l}\frac{d^2\mu_l}{d\eta_l^2}, \\
g_l = g(\mu_l) &= \frac{1}{V_l}\frac{d\mu_l}{d\eta_l}\frac{d^2\mu_l}{d\eta_l^2} - \frac{1}{V_l^2}\frac{dV_l}{d\mu_l}\left(\frac{d\mu_l}{d\eta_l}\right)^3, \\
b_l = b(\mu_l) &= \frac{1}{V_l^3}\left(\frac{d\mu_l}{d\eta_l}\right)^4 \left\{\left(\frac{dV_l}{d\mu_l}\right)^2 + V_l\frac{d^2V_l}{d\mu_l^2}\right\},
\end{aligned}$$

and

$$h_l = h(\mu_l) = \frac{1}{V_l^2}\frac{dV_l}{d\mu_l}\left(\frac{d\mu_l}{d\eta_l}\right)^2\frac{d^2\mu_l}{d\eta_l^2} + \frac{1}{V_l^2}\frac{d^2V_l}{d\mu_l^2}\left(\frac{d\mu_l}{d\eta_l}\right)^4,$$

for $l = 1, \ldots, n$. The A's can be written as

$$A_1 = A_{1,\beta} + A_{1,\beta\phi}, \quad A_2 = A_{2,\beta} + A_{2,\beta\phi}, \quad A_3 = A_{3,\beta},$$

where

$$
\begin{aligned}
A_{1,\beta} = \phi^{-1} \Big[& 3\mathbf{1}^\top \boldsymbol{F}\boldsymbol{Z}_{2d}(\boldsymbol{Z}-\boldsymbol{Z}_2)\boldsymbol{Z}_{2d}\boldsymbol{F}\mathbf{1} + 6\mathbf{1}^\top \boldsymbol{F}\boldsymbol{Z}_{2d}\boldsymbol{Z}_2(\boldsymbol{Z}-\boldsymbol{Z}_2)_d(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} \\
& -6\mathbf{1}^\top \boldsymbol{F}\{\boldsymbol{Z}_2^{(2)} * (\boldsymbol{Z}-\boldsymbol{Z}_2)\}(2\boldsymbol{G}-\boldsymbol{F})\mathbf{1} - 6\mathbf{1}^\top \boldsymbol{H}(\boldsymbol{Z}-\boldsymbol{Z}_2)_d\boldsymbol{Z}_{2d}\mathbf{1} \Big], \\
A_{2,\beta} = \phi^{-1} \Big[& -3\mathbf{1}^\top (\boldsymbol{F}-\boldsymbol{G})(\boldsymbol{Z}-\boldsymbol{Z}_2)_d\boldsymbol{Z}_2(\boldsymbol{Z}-\boldsymbol{Z}_2)_d(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} \\
& - 6\mathbf{1}^\top \boldsymbol{F}\boldsymbol{Z}_{2d}(\boldsymbol{Z}-\boldsymbol{Z}_2)(\boldsymbol{Z}-\boldsymbol{Z}_2)_d(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} \\
& -6\mathbf{1}^\top (\boldsymbol{F}-\boldsymbol{G})\{(\boldsymbol{Z}-\boldsymbol{Z}_2)^{(2)} * \boldsymbol{Z}_2\}(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} + 3\mathbf{1}^\top \boldsymbol{B}(\boldsymbol{Z}-\boldsymbol{Z}_2)_d^{(2)}\mathbf{1} \Big],
\end{aligned}
$$

and

$$
\begin{aligned}
A_{3,\beta} = \phi^{-1} \Big[& 3\mathbf{1}^\top (\boldsymbol{F}-\boldsymbol{G})(\boldsymbol{Z}-\boldsymbol{Z}_2)_d(\boldsymbol{Z}-\boldsymbol{Z}_2)(\boldsymbol{Z}-\boldsymbol{Z}_2)_d(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} \\
& +2\mathbf{1}^\top (\boldsymbol{F}-\boldsymbol{G})(\boldsymbol{Z}-\boldsymbol{Z}_2)^{(3)}(\boldsymbol{F}-\boldsymbol{G})\mathbf{1} \Big].
\end{aligned}
$$

Here "$*$" represents the Hadamard product (elementwise product), $\boldsymbol{Z}^{(2)} = \boldsymbol{Z} * \boldsymbol{Z}$, $\boldsymbol{Z}^{(3)} = \boldsymbol{Z}^{(2)} * \boldsymbol{Z}$ and $\boldsymbol{Z}_d = \mathrm{diag}(z_{11}, \ldots, z_{nn})$, with z_{ll} denoting the lth diagonal element of $\boldsymbol{Z}$. When ϕ is known, $A_{1,\beta\phi}$ and $A_{2,\beta\phi}$ are equal to zero (Cordeiro *et al.* 1993). For unknown ϕ, we assume that the term $a(y,\phi)$ in (1) can be written as $a(y,\phi) = d_1(\phi) + d_2(y)$. Let $d_{(2)} = d_{(2)}(\phi) = \phi^2 d_1''(\phi)$ and $d_{(3)} = d_{(3)}(\phi) = \phi^3 d_1'''(\phi)$. Cribari-Neto and Ferrari (1995) showed that

$$
A_{1,\beta\phi} = \frac{6q\{d_{(3)} + (2-p+q)d_{(2)}\}}{nd_{(2)}^2}
$$

and

$$
A_{2,\beta\phi} = \frac{3q(q+2)}{nd_{(2)}}.
$$

Affiliation:

Antonio Hermes M. da Silva-Júnior, Damião Nóbrega da Silva
Departamento de Estatística
Universidade Federal do Rio G. do Norte
Campus Universitário s/n
Natal, RN, 59078–970, Brazil
E-mail: `hermes@ccet.ufrn.br`, `damiao@ccet.ufrn.br`

Silvia L. P. Ferrari
Departamento de Estatística
Universidade de São Paulo
Rua do Matão, 1010
São Paulo, SP, 05508-090, Brazil
E-mail: `silviaferrari.usp@gmail.com`

Journal of Statistical Software `http://www.jstatsoft.org/`
published by the American Statistical Association `http://www.amstat.org/`
Volume 61, Code Snippet 2 *Submitted:* 2011-12-16
October 2014 *Accepted:* 2014-06-02