Computational Statistics and Data Analysis 53 (2009) 3397–3411

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

The distribution of Pearson residuals in generalized linear models

Gauss M. Cordeiro [a,*], Alexandre B. Simas [b]

[a] Depto de Estatística e Informática, Univers. Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros s/n, Dois Irmãos, 52171-900, Recife, Brazil

[b] Associação Instituto Nacional de Matemática Pura e Aplicada, Estrada D. Castorina, 110, Jd. Botânico, 22460-320, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 9 May 2008
Received in revised form 31 December 2008
Accepted 14 February 2009
Available online 9 March 2009

ABSTRACT

In general, the distribution of residuals cannot be obtained explicitly. In this paper we give an asymptotic formula for the density of Pearson residuals in continuous generalized linear models corrected to order n^{-1}, where n is the sample size. We define a set of corrected Pearson residuals for these models that, to this order of approximation, have exactly the same distribution of the true Pearson residuals. An application to a real data set and simulation results for a gamma model illustrate the usefulness of our corrected Pearson residuals.

1. Introduction

The residuals carry important information concerning the appropriateness of assumptions that underlie statistical models, and thereby play an important role in checking model adequacy. They are used to identify discrepancies between models and data, so it is natural to base residuals on the contributions made by individual observations to measures of model fit. The use of residuals for assessing the adequacy of fitted regression models is nowadays commonplace due to the widespread availability of statistical software, many of which are capable of displaying residuals and diagnostic plots, at least for the more commonly used models. Beyond special models, relatively little is known about asymptotic properties of residuals in general regression models. There is a clear need to study second-order asymptotic properties of appropriate residuals to be used for diagnostic purposes in nonlinear regression models.

The paper by Nelder and Wedderburn (1972) first identified and unified the theory of generalized linear models (GLMs), including a general algorithm for computing maximum likelihood estimates (MLEs). In continuous GLMs, the random variables $Y_1, \ldots, Y_n$ are assumed independent and each Y_i has a density function in the linear exponential family

$$\pi(y; \theta_i, \phi) = \exp[\phi\{y\theta_i - b(\theta_i)\} + c(y, \phi)], \tag{1}$$

where $b(\cdot)$ and $c(\cdot, \cdot)$ are known appropriate functions. We assume that the precision parameter $\phi = \sigma^{-2}$, σ^2 is the so-called dispersion parameter, is the same for all observations, although possibly unknown. We also assume a probability density function π with respect to the Lebesgue measure. We do not consider the discrete distributions such as Poisson, binomial and negative binomial which take the form (1). For two-parameter full exponential family distributions with canonical parameters ϕ and $\phi\theta$, the decomposition $c(y, \phi) = \phi a(y) + d_1(y) + d_2(\phi)$ holds. The mean and variance of Y_i are, respectively, $E(Y_i) = \mu_i = \mathrm{d}b(\theta_i)/\mathrm{d}\theta_i$ and $\mathrm{Var}(Y_i) = \phi^{-1}V_i$, where $V = \mathrm{d}\mu/\mathrm{d}\theta$ is the variance function. For gamma models, the dispersion parameter σ^2 is the reciprocal of the index; for normal and inverse Gaussian models, σ^2 is the variance and $\mathrm{Var}(Y_i)/E(Y_i)^3$, respectively. The parameter $\theta = \int V^{-1}\mathrm{d}\mu = q(\mu)$ is a known one-to-one function of μ. A linear exponential family is characterized by its variance function, which plays a key role in estimation.

* Corresponding author. Tel.: +55 81 99317558.
E-mail addresses: gausscordeiro@uol.com.br (G.M. Cordeiro), alesimas@impa.br (A.B. Simas).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.02.025

A GLM is defined by the family of distributions (1) and the systematic component $g(\mu) = \eta = X\beta$, where $g(\cdot)$ is a known one-to-one continuously twice-differentiable function, X is a specified $n \times p$ model matrix of full rank $p < n$ and $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a set of unknown linear parameters to be estimated. Let $\hat{\beta}$ be the MLE of β.

Residuals in GLMs were first discussed by Pregibon (1981), though ostensibly concerned with logistic regression models, Williams (1984, 1987) and Pierce and Schafer (1986). McCullagh and Nelder (1989) provided a survey of GLMs, with substantial attention to definition of residuals. Pearson residuals are the most commonly used measure of overall fit for GLMs. They are defined by $R_i = (Y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$, where $\hat{\mu}_i$ and $\hat{V}_i$ are respectively the fitted mean and fitted variance function of Y_i. In this paper we consider only Pearson residuals appropriate to our particular asymptotic aims when $n \to \infty$. Cordeiro (2004) obtained matrix formulae for the expectations, variances and covariances of these residuals, and defined adjusted Pearson residuals in these models having zero mean and unit variance to order n^{-1}. The Pearson residuals defined in Cordeiro (2004) are proportional to $\sqrt{\phi}$, although we are considering here R_i as usual without the precision parameter ϕ. While Cordeiro's adjusted Pearson residuals do correct the residuals for equal mean and variance, the distribution of these residuals is not equal to the distribution of the true Pearson residuals to order n^{-1}. Further, Cordeiro and Paula (1989) introduced the class of exponential family nonlinear models which generalizes the GLMs. Later, Wei (1998) gave a comprehensive introduction to exponential family nonlinear models. More recently, Simas and Cordeiro (2009) generalized Cordeiro's (2004) result by obtaining matrix formulae for the $\mathcal{O}(n^{-1})$ expectations, variances and covariances for Pearson residuals in exponential family nonlinear models.

In a general setup, the distribution of residuals usually differs from the distribution of the true residuals by terms of order n^{-1}. Cox and Snell (1968) discussed a general definition of residuals, applicable to a wide range of models, and obtained useful expressions to order $\mathcal{O}(n^{-1})$ for their first two moments, where n is the sample size. Loynes (1969) derived, under some regularity conditions, and again to order n^{-1}, the asymptotic expansion for the density function of the Cox and Snell's residuals, and then defined corrected residuals having the same distribution as the random variables which they are effectively estimating. In all but the simplest situations, the use of the results by Cox, Snell and Loynes will require a considerable amount of tedious algebra. Our chief goal is to obtain an explicit formula for the density of the Pearson residuals to order n^{-1} which holds for all continuous GLMs. In Section 2 we give a summary of key results from Loynes (1969) applied to Pearson residuals in GLMs. The density of Pearson residuals in these models corrected to order n^{-1} is presented in Section 3. In Section 4 we provide applications to some common models. In Section 5 we compare our corrected residuals with the adjusted residuals proposed by Cordeiro (2004). In Section 6, we present simulation studies to assess the adequacy of the approximations for a gamma model with log link. In Section 7, we provide an application to a real data set that demonstrates the usefulness of the corrected residuals. Concluding remarks are given in Section 8 and, in the Appendix, we present a more rigorous proof of the general results discussed by Loynes (1969).

2. Conditional moments of Pearson residuals

Let l_i be the log-likelihood contribution from Y_i. We obtain from (1)

$$l_i = \phi\,\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)$$

and then the ith element of the score function is simply

$$U_r^{(i)} = \frac{\partial l_i}{\partial \beta_r} = \phi V_i^{-1/2} w_i^{1/2}(Y_i - \mu_i)x_{ir},$$

where $w = V^{-1}\mu'^2$ is the weight function and from now on dashes indicate derivatives with respect to η. Let $\varepsilon_i = V_i^{-1/2}(Y_i - \mu_i)$ be the true Pearson residual corresponding to the Pearson residual $R_i = \hat{V}_i^{-1/2}(Y_i - \hat{\mu}_i)$. Suppose we write the Pearson residual as $R_i = \varepsilon_i + \delta_i$. We can write the following conditional moments given $\varepsilon_i = x$ to order n^{-1} (Loynes, 1969)

$$\mathrm{Cov}(\hat{\beta}_r, \hat{\beta}_s \mid \varepsilon_i = x) = -\kappa^{rs},$$

$$b_s^{(i)}(x) = E(\hat{\beta}_s - \beta_s \mid \varepsilon_i = x) = B(\hat{\beta}_s) - \sum_{r=1}^{p} \kappa^{sr} U_r^{(i)}(x), \tag{2}$$

where $-\kappa^{sr}$ is the (s, r)th element of the inverse Fisher information matrix K^{-1} for β, $B(\hat{\beta}_s)$ is the $\mathcal{O}(n^{-1})$ bias of $\hat{\beta}_s$ and $U_r^{(i)}(x) = E(U_r^{(i)} \mid \varepsilon_i = x)$ is the conditioned score function. The mean and variance of the asymptotic distribution of δ_i, given $\varepsilon_i = x$, are respectively to order n^{-1}

$$\theta_x^{(i)} = E(\delta_i \mid \varepsilon_i = x) = \sum_{r=1}^{p} H_r^{(i)}(x) b_r^{(i)}(x) - \frac{1}{2}\sum_{r,s}^{p} H_{rs}^{(i)}(x)\kappa^{rs}, \tag{3}$$

$$\phi_x^{(i)^2} = \mathrm{Var}(\delta_i \mid \varepsilon_i = x) = -\sum_{r,s=1}^{p} H_r^{(i)}(x) H_s^{(i)}(x)\kappa^{rs}, \tag{4}$$

where $H_r^{(i)} = \partial\varepsilon_i/\partial\beta_r$, $H_{rs}^{(i)} = \partial^2\varepsilon_i/\partial\beta_r\partial\beta_s$, $H_r^{(i)}(x) = E(H_r^{(i)} \mid \varepsilon_i = x)$ and $H_{rs}^{(i)}(x) = E(H_{rs}^{(i)} \mid \varepsilon_i = x)$. Let $V_i^{(m)} = \frac{\mathrm{d}^m V_i}{\mathrm{d}\mu_i^m}$ for $m = 1, 2$. We obtain by simple differentiation

$$H_r^{(i)} = \{-V_i^{-1/2}\mu_i' - \frac{1}{2}V_i^{-3/2}V_i^{(1)}\mu_i'(Y_i - \mu_i)\}\, x_{ir}$$

and

$$\begin{aligned} H_{rs}^{(i)} = \{&-V_i^{-1/2}\mu_i'' + V_i^{-3/2}V_i^{(1)}\mu_i'^2 + \frac{3}{4}V_i^{-5/2}V_i^{(1)^2}\mu'^2(Y_i - \mu_i) \\ &- \frac{1}{2}V_i^{-3/2}V_i^{(2)}\mu_i'^2(Y_i - \mu_i) - \frac{1}{2}V_i^{-3/2}V_i^{(1)}\mu_i'(Y_i - \mu_i)\}\, x_{ir}x_{is}. \end{aligned}$$

Conditioning on $\varepsilon_i = x$ leads to $H_r^{(i)}(x) = e_i(x)x_{ir}$ and $H_{rs}^{(i)}(x) = h_i(x)x_{ir}x_{is}$, where

$$e_i(x) = -V_i^{-1/2}\mu_i' - \frac{1}{2}V_i^{-1}V_i^{(1)}\mu_i'\, x \tag{5}$$

and

$$h_i(x) = -V_i^{-1/2}\mu_i'' + V_i^{-3/2}V_i^{(1)}\mu_i'^2 + \frac{1}{4}\{(3V_i^{-2}V_i^{(1)^2} - 2V_i^{-1}V_i^{(2)})\mu_i'^2 - 2V_i^{-1}V_i^{(1)}\mu_i''\}\, x. \tag{6}$$

For canonical models ($\theta = \eta$), (5) and (6) become

$$e_i(x) = -V_i^{1/2} - \frac{V_i^{(1)}}{2}x \quad\text{and}\quad h_i(x) = \frac{1}{4}(V_i^{(1)^2} - 2V_iV_i^{(2)})x.$$

Conditioning the score function $U_r^{(i)} = \phi V_i^{-1/2}w_i^{1/2}(Y_i - \mu_i)x_{ir}$ on $\varepsilon_i = x$ yields $U_r^{(i)}(x) = \phi\, w_i^{1/2}\, x_{ir}\, x$, and then using (2) we find

$$b_s^{(i)}(x) = B(\hat\beta_s) + \phi\, w_i^{1/2}\, x\, \tau_s^{\mathrm{T}}K^{-1}X^{\mathrm{T}}\gamma_i,$$

where $K^{-1} = \phi^{-1}(X^{\mathrm{T}}WX)^{-1}$, $W = \mathrm{diag}\{w_i\}$ is a diagonal matrix of weights, τ_s is a p-vector with one in the *sth* position and zeros elsewhere and γ_i is an n-vector with one in the *ith* position and zeros elsewhere. Defining $M = \{m_{si}\} = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}$, it is easily verified that

$$b_s^{(i)}(x) = w_i^{1/2}\, m_{si}\, x + B(\hat\beta_s).$$

Cordeiro and McCullagh (1991) showed that the n^{-1} bias of $\hat\beta$ is given by

$$B(\hat\beta) = -(2\phi)^{-1}(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Z_d\, F\, 1,$$

where $F = \mathrm{diag}\{V_i^{-1}\mu_i'\,\mu_i''\}$, $Z = \{z_{ij}\} = X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}$, $Z_d = \mathrm{diag}\{z_{ii}\}$ is a diagonal matrix with the diagonal elements of Z and 1 is an n-vector of ones. The asymptotic covariance matrix of the MLE $\hat\eta$ of the linear predictor is simply $\phi^{-1}Z$. We obtain

$$\begin{aligned} \sum_{r=1}^{p} H_r^{(i)}(x)b_r^{(1)}(x) &= e_i(x)\left\{x\, w_i^{1/2}\sum_{r=1}^{p} m_{ri}x_{ir} + \sum_{r=1}^{p} B(\hat\beta_r)x_{ir}\right\} \\ &= e_i(x)\{w_i^{1/2}z_{ii}\, x + B(\hat\eta_i)\}, \end{aligned}$$

where $B(\hat\eta_i)$ is the ith element of the $\mathcal{O}(n^{-1})$ bias $B(\hat\eta) = -(2\phi)^{-1}ZZ_d\, F\, 1$ of $\hat\eta$. This bias depends on the model matrix, the variance function and the first two derivatives of the link function. Also,

$$-\frac{1}{2}\sum_{r,s=1}^{p} H_{rs}^{(i)}(x)\kappa^{rs} = -\,\frac{z_{ii}}{2\phi}h_i(x).$$

The conditional mean $\theta_x^{(i)}$ from (3) is then a second-degree polynomial in x given by

$$\theta_x^{(i)} = \{w_i^{1/2}z_{ii}x + B(\hat\eta_i)\}e_i(x) + \frac{z_{ii}}{2\phi}h_i(x), \tag{7}$$

where $e_i(x)$ and $h_i(x)$ are obtained from (5) and (6).

We now compute the conditional variance $\phi_x^{(i)^2}$. From (4) it follows

$$\phi_x^{(i)^2} = \frac{z_{ii}}{\phi}e_i(x)^2. \tag{8}$$

Hence, $\phi_x^{(i)^2}$ is also a second-degree polynomial in x.

Table 1
Densities of the true residuals for some distributions.

Distribution	Density in (1)	Density of the true residual ($f_\varepsilon(x)$)
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right), x\in\mathbb{R}$
Gamma	$\frac{(\phi x)^{\phi-1}\phi}{\Gamma(\phi)\mu^\phi}\exp(-\phi x/\mu)$	$\frac{\{\phi(1+x)\}^{\phi-1}\phi}{\Gamma(\phi)}\exp\{-\phi(1+x)\}, x>-1$
Inverse Gaussian	$\frac{\sqrt{\phi}}{\sqrt{2\pi x^3}}\exp\left\{-\frac{\phi(x-\mu)^2}{2\mu^2 x}\right\}$	$\left\{\frac{\phi}{2\pi(\mu^{1/2}x+1)^3}\right\}^{\frac{1}{2}}\exp\left\{-\frac{\phi x^2}{2(\mu^{1/2}x+1)}\right\}, x>\frac{-1}{\sqrt{\mu}}$

3. The density of Pearson residuals

A simple calculation from (1) gives the density of the true Pearson residual

$$f_{\varepsilon_i}(x)=\sqrt{V_i}\,\exp\left[\phi\left\{\sqrt{V_i}\,\theta_i x+\mu_i\theta_i-b(\theta_i)\right\}+c\left(\sqrt{V_i}\,x+\mu_i,\phi\right)\right], \tag{9}$$

where $\theta=q(\mu)$. Table 1 gives the densities of the true Pearson residuals for the normal, gamma and inverse Gaussian distributions, where $\Gamma(\cdot)$ is the gamma function.

Throughout the following we shall assume all necessary regularity conditions are satisfied. The density function of the Pearson residual R_i in continuous GLMs to order n^{-1} follows from Loynes (1969). See, also, equation (21). We have

$$f_{R_i}(x)=f_{\varepsilon_i}(x)-\frac{\mathrm{d}\{f_{\varepsilon_i}(x)\theta_x^{(i)}\}}{\mathrm{d}x}+\frac{1}{2}\frac{\mathrm{d}^2\{f_{\varepsilon_i}(x)\phi_x^{(i)^2}\}}{\mathrm{d}x^2}, \tag{10}$$

where $f_{\varepsilon_i}(x)$, $\theta_x^{(i)}$ and $\phi_x^{(i)^2}$ come from (9), (7) and (8), respectively.

We now define corrected Pearson residuals for GLMs of the form $R_i'=R_i+\rho_i(R_i)$, where $\rho(\cdot)$ is a function of order $\mathcal{O}(n^{-1})$ constructed in order to produce the residual R_i' with the same distribution of ε_i to order n^{-1}. Loynes (1969) showed (see, also, the proof given in the Appendix) that

$$\rho_i(x)=-\theta_x^{(i)}+\frac{1}{2f_{\varepsilon_i}(x)}\frac{\mathrm{d}\{f_{\varepsilon_i}(x)\phi_x^{(i)^2}\}}{\mathrm{d}x} \tag{11}$$

makes $f_{R_i'}(x)=f_{\varepsilon_i}(x)$ to order n^{-1}, i.e., the corrected residuals R_i' have exactly the same distribution of the true residuals to this order. Combining (8) with (9) gives

$$\frac{1}{2f_{\varepsilon_i}(x)}\frac{\mathrm{d}\{f_{\varepsilon_i}(x)\phi_x^{(i)^2}\}}{\mathrm{d}x}=\frac{z_{ii}}{\phi}e_i(x)\frac{\mathrm{d}e_i(x)}{\mathrm{d}x}+\frac{z_{ii}}{2\phi}e_i(x)^2\left\{\phi\sqrt{V_i}\,\theta_i+\frac{\mathrm{d}}{\mathrm{d}x}c(\sqrt{V_i}\,x+\mu_i,\phi)\right\}. \tag{12}$$

Using (11), (7) and (12), the correction function turns out to be

$$\begin{aligned}\rho_i(x)&=e_i(x)\left\{-\frac{1}{2\phi}V_i^{-1}V_i^{(1)}\mu_i'z_{ii}-B(\hat{\eta}_i)-w_i^{1/2}z_{ii}\,x\right\}\\&\quad-\frac{z_{ii}}{2\phi}h_i(x)+\frac{z_{ii}}{2\phi}e_i(x)^2\left\{\phi\sqrt{V_i}\,q(\mu_i)+\frac{\mathrm{d}}{\mathrm{d}x}c\left(\sqrt{V_i}x+\mu_i,\phi\right)\right\}.\end{aligned} \tag{13}$$

Direct substitution using (13) yields corrected Pearson residuals R_i' for most GLMs. The term $\phi^{-1}z_{ii}$ in this equation is just $\mathrm{Var}(\hat{\eta}_i)$. Although there are several terms in (13), this expression is simply applied to any continuous model since we need only to calculate $e_i(x)$, $h_i(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}c(\sqrt{V_i}x+\mu_i,\phi)$ from (5), (6) and (1), the others terms being standard quantities in the theory of GLMs. More generally, the corrected residuals R_i' depend on the GLM only through the model matrix X, the precision parameter ϕ, the function $c(\cdot,\cdot)$ and the variance and link functions with their first two derivatives.

The density of the true residual for the inverse Gaussian model given in Table 1 depends on the unknown mean μ. However, we can estimate this density using the general expression for the corrected MLE of μ, $\tilde{\mu}$ say, given by Cordeiro and McCullagh (1991), formula (4.4). The resulting estimated density is identical to the true density except by terms of order less than n^{-1} and the results of Sections 3 and 4 could also be applied to this distribution. To prove this, let $\tilde{\mu}=\mu+c/n^2$. Then, keeping only terms up to order n^{-2}, we have

$$\tilde{\mu}^{1/2}=\sqrt{\mu}\sqrt{1+\frac{c}{n^2\mu}}=\sqrt{\mu}\left(1+\frac{c}{2n^2\mu}\right).$$

Table 2
Values of μ', μ'' and w for some link functions.

Link function	Formula	μ'	μ''	w
Linear	$\mu = \eta$	1	0	V^{-1}
Log	$\log(\mu) = \eta$	μ	μ	$\mu^2 V^{-1}$
Reciprocal	$\mu^{-1} = \eta$	$-\mu^2$	$2\mu^3$	$\mu^4 V^{-1}$
Inverse of the square	$\mu^{-2} = \eta$	$-\mu^3/2$	$3\mu^5/4$	$\mu^6 V^{-1}/4$

Table 3
Values of $q(\mu)$, V, w, and $\frac{\mathrm{d}}{\mathrm{d}x} c(\sqrt{V}x + \mu, \phi)$ for the normal, gamma and inverse Gaussian distributions.

Model	$q(\mu)$	V	w	$\frac{\mathrm{d}}{\mathrm{d}x} c(\sqrt{V}x + \mu, \phi)$
Normal	μ	1	μ'^2	$-(x+\mu)\phi$
Gamma	$-1/\mu$	μ^2	$\mu^{-2}\mu'^2$	$(\phi - 1)/(1+x)$
Inverse Gaussian	$-1/(2\mu^2)$	μ^3	$\mu^{-3}\mu'^2$	$-\frac{3\mu^{3/2}}{2(\mu^{3/2}x+\mu)} + \frac{\phi\mu^{3/2}}{2(\mu^{3/2}x+\mu)^2}$

Also,

$$(\tilde{\mu}^{1/2}x + 1)^{-3/2} = (\sqrt{\mu}x + 1)^{-3/2}\left\{1 - \frac{3\,x\,c}{4n^2\sqrt{\mu}(\sqrt{\mu}x + 1)}\right\}$$

and

$$\exp\left\{\frac{-\phi x^2}{2(\tilde{\mu}^{1/2}x + 1)}\right\} = \exp\left[\frac{-\phi x^2}{2(\sqrt{\mu}x + 1)}\frac{1}{\{1 + \frac{x\,c}{2n^2\sqrt{\mu}(\sqrt{\mu}x+1)}\}}\right].$$

Then,

$$\exp\left\{\frac{-\phi x^2}{2(\tilde{\mu}^{1/2}x + 1)}\right\} = \exp\left\{\frac{-\phi\, x^2}{2(\sqrt{\mu}x + 1)}\right\}\exp\left\{\frac{\phi\, x^3\, c}{4n^2\sqrt{\mu}(\sqrt{\mu}x + 1)^2}\right\}.$$

Hence,

$$\sqrt{\frac{\phi}{2\pi}}\frac{1}{(\tilde{\mu}^{1/2}x + 1)^{3/2}}\exp\left\{\frac{-\phi x^2}{2(\tilde{\mu}^{1/2}x + 1)}\right\} = \sqrt{\frac{\phi}{2\pi}}(\sqrt{\mu}x + 1)^{-3/2}\left(1 - \frac{c_1}{n^2}\right)\exp\left\{\frac{-\phi x^2}{2(\sqrt{\mu}x + 1)}\right\}\exp\left(\frac{c_2}{n^2}\right),$$

where $c_1 = \frac{3\,x\,c}{4\sqrt{\mu}(\sqrt{\mu}x+1)}$ and $c_2 = \frac{\phi\, x^3\, c}{4\sqrt{\mu}(\sqrt{\mu}x+1)^2}$. This equation shows that the estimated density and the true density of ε agree to order n^{-1}.

The results of this section have been obtained assuming that the dispersion parameter is known. However, we show in Sections 6 and 7 that they can be applied even when the dispersion parameter is replaced by a consistent estimate.

4. Some special models

Formula (13) holds for all continuous GLMs including the models in common use: linear models, canonical models, normal models, gamma models and inverse Gaussian models. In this section we will compute the correction $\rho_i(\cdot)$ in (13) for some important GLMs from which we can obtain the corrected residuals $R_i' = R_i + \rho_i(R_i)$. Table 2 gives the values of μ', μ'' and w for some useful link functions and Table 3 gives $q(\mu)$, V, w and $\frac{\mathrm{d}}{\mathrm{d}x} c(\sqrt{V}x + \mu, \phi)$ for the normal, gamma and inverse Gaussian distributions.

4.1. Linear models

For linear models, $\mu_i = \eta_i$, $\mu_i' = 1$, $\mu_i'' = 0$, $w_i = V_i^{-1}$, $B(\hat{\eta}_i) = 0$ and then $e_i(x) = -V_i^{-1/2} - \frac{1}{2}V_i^{-1}V_i^{(1)}x$ and $h_i(x) = V_i^{-3/2}V_i^{(1)} + \frac{3}{4}V_i^{-2}V_i^{(1)^2}x - \frac{1}{2}V_i^{-1}V_i^{(2)}x$. Thus, we have

$$\rho_i(x) = V_i^{-1}z_{ii}\,x\left(1 - \frac{V_i^{-1}V_i^{(1)^2}}{8\phi} + \frac{V_i^{(2)}}{4\phi} + \frac{V_i^{-1/2}V_i^{(1)}}{2}x\right)$$
$$+ \frac{z_{ii}}{2\phi}\left(V_i^{-1} + V_i^{-3/2}V_i^{(1)}\,x + \frac{1}{4}V_i^{-2}V_i^{(1)^2}\,x^2\right)\left\{\phi\sqrt{V_i}q(\mu_i) + \frac{\mathrm{d}}{\mathrm{d}x}c(\sqrt{V_i}\,x + \mu_i, \phi)\right\}.$$

4.2. Canonical models

For canonical models, $\eta_i = \theta_i$, $w_i = V_i$, $\mu_i' = V_i$ and $\mu_i'' = V_i V_i^{(1)}$. Further, $e_i(x) = -V_i^{1/2} - \frac{1}{2}V_i^{(1)}x$ and $h_i(x) = \frac{1}{4}(V_i^{(1)^2} - 2V_iV_i^{(2)})\,x$. Hence,

$$\rho_i(x) = \left(V_i^{1/2} + \frac{V_i^{(1)}}{2}x\right)B(\hat{\eta}_i) + z_{ii}\left(\frac{V_i^{1/2}V_i^{(1)}}{2\phi} + V_i x + \frac{V_i^{(1)^2}}{8\phi}x + \frac{V_iV_i^{(2)}}{4\phi}x\right) + \frac{V_i^{1/2}V_i^{(1)}}{2}x^2$$
$$+ \frac{z_{ii}}{2\phi}\left(V_i + V_i^{1/2}V_i^{(1)}x + \frac{1}{4}V_i^{(1)^2}x^2\right)\left\{\phi\sqrt{V_i}\,q(\mu_i) + \frac{\mathrm{d}}{\mathrm{d}x}c\left(\sqrt{V_i}\,x + \mu_i, \phi\right)\right\}.$$

4.3. Normal models

For normal models, $V_i = 1$, $w_i = \mu_i'^2$, $c(x,\phi) = -\{x^2\phi + \log(2\pi/\phi)\}/2$, $\frac{\mathrm{d}}{\mathrm{d}x}c(x+\mu,\phi) = -(x+\mu)\phi$, $e_i(x) = -\mu_i'$ and $h_i(x) = -\mu_i''$. We have

$$\rho_i(x) = B(\hat{\eta}_i)\mu_i' + \frac{\mu_i''\,z_{ii}}{2\phi} + \frac{\mu_i'^2 z_{ii}}{2}\,x.$$

As a special case of the normal model, we consider the normal linear model for which $\mu = \theta = \eta$, $e_i(x) = -1$ and $h_i(x) = 0$. Then, we obtain

$$\rho_i(x) = z_{ii}\,x/2,$$

and the corrected residuals become

$$R_i' = R_i\left(1 + \frac{z_{ii}}{2}\right).$$

It is easily proved that $\mathrm{Var}(R_i') = 1 + \mathcal{O}(n^{-2})$. A check of this expression follows for the simplest case of independent and identically distributed observations. We have $Z = n^{-1}\,1\,1^{\mathrm{T}}$, $z_{ii} = n^{-1}$ and then

$$R_i' = R_i\left(1 + \frac{1}{2n}\right),$$

which is identical to the equation given in the example discussed by Loynes (1969).

4.4. Gamma models

For gamma models, $V_i = \mu_i^2$, $w_i = \mu_i^{-2}\mu_i'^2$, $c(x,\phi) = (\phi - 1)\log(x) + \phi\log(\phi) - \log\Gamma(\phi)$ and $\frac{\mathrm{d}}{\mathrm{d}x}c(\mu x + \mu, \phi) = (\phi-1)/(1+x)$. We have $e_i(x) = -\mu_i^{-1}\mu_i' - \mu_i^{-1}\mu_i' x$ and $h_i(x) = -\mu_i^{-1}\mu_i'' + 2\mu_i^{-2}\mu_i'^2 - \mu_i^{-1}\mu_i''\,x + 2\mu_i^{-2}\mu_i'^2\,x$. Then,

$$\rho_i(x) = (1+x)\left(\mu_i^{-1}\mu_i' B(\hat{\eta}_i) + \frac{\mu_i^{-1}\mu_i''}{2\phi}z_{ii} - \frac{\mu_i^{-2}\mu_i'^2}{2\phi}z_{ii} + \frac{\mu_i^{-2}\mu_i'^2 z_{ii}}{2}x\right).$$

4.5. Inverse Gaussian models

For inverse Gaussian models, $V_i = \mu_i^3$, $w_i = \mu_i^{-3}\mu_i'^2$, $c(x,\phi) = (1/2)\log\{\phi/(2\pi x^3)\} - \phi/(2x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}c(\mu^{3/2}x + \mu, \phi) = -\frac{3\mu^{3/2}}{2(\mu^{3/2}x+\mu)} + \frac{\phi\mu^{3/2}}{2(\mu^{3/2}x+\mu)^2}$. Further, $e_i(x) = -\mu_i^{-3/2}\mu_i' - \frac{3}{2}\mu_i^{-1}\mu_i' x$ and $h_i(x) = -\mu_i^{-3/2}\mu_i'' + 3\mu_i^{-5/2}\mu_i'^2 + \frac{15}{4}\mu_i^{-2}\mu_i'^2 x - \frac{3}{2}\mu_i^{-1}\mu_i'' x$. Then,

$$\rho_i(x) = \left(\mu_i^{-3/2}\mu_i' + \frac{3\mu_i'}{2\mu_i}x\right)B(\hat{\eta}_i) + \frac{\mu_i^{3/2}\mu_i''z_{ii}}{2\phi} + \left(\frac{3\mu_i'^2 z_{ii}}{8\phi\mu_i^2} + \frac{3\mu_i''z_{ii}}{4\phi\mu_i}\right)x + \frac{\mu_i'^2 z_{ii}}{\mu_i^3}x + \frac{3\mu_i'^2 z_{ii}}{2\mu_i^{5/2}}\,x^2$$
$$+ \frac{\mu_i^{-3}z_{ii}}{4\phi}\left(\mu_i'^2 + 3\mu_i^{1/2}\mu_i'^2 x + \frac{9\mu_i\mu_i'^2}{4}x^2\right)\left\{-\frac{\phi}{\mu_i^{1/2}} - \frac{3\mu_i^{3/2}}{(\mu_i^{3/2}x + \mu_i)} + \frac{\phi\mu_i^{3/2}}{(\mu_i^{3/2}x+\mu_i)^2}\right\}.$$

5. Expansion for Cordeiro's adjusted residual

In this section we derive the density function of the adjusted Pearson residuals proposed by Cordeiro (2004). He gave simple expressions to order n^{-1} for the mean and variance of the Pearson residual R_i in GLMs, namely $E(R_i) = m_i/n + \mathcal{O}(n^{-2})$ and $\mathrm{Var}(R_i) = \sigma^2 + v_i/n + \mathcal{O}(n^{-2})$, where

$$\frac{m_i}{n} = -\frac{\sigma^2}{2}\gamma_i(I - H)Jz \quad \text{and} \quad \frac{v_i}{n} = \frac{\sigma^4}{2}\gamma_i(QHJ - T)z,$$

I is the identity matrix of order n, $H = W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2}$ is the projection matrix, J, Q and T are diagonal matrices given by $J = \mathrm{diag}\{V_i^{-1/2}\mu_i''\}$, $Q = \mathrm{diag}\{V_i^{-1/2}V_i^{(1)}\}$, $T = \mathrm{diag}\{2\phi w_i + w_i V_i^{(2)} + V_i^{-1}V_i^{(1)}\mu_i''\}$, $z = (z_{11}, \ldots, z_{nn})^{\mathrm{T}}$ is an n-vector with the diagonal elements of $Z = X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}$, and γ_i is, as defined in Section 2, an n-vector with one in the ith position and zeros elsewhere. The adjusted residuals proposed by Cordeiro (2004) are

$$R_i^* = \frac{R_i - \hat{m}_i/n}{(\sigma^2 + \hat{v}_i/n)^{1/2}}. \tag{14}$$

Expanding $(\sigma^2 + \frac{\hat{v}_i}{n})^{-1/2}$ as $\sigma^{-1}(1 - \frac{\hat{v}_i}{2n\sigma^2} + \cdots)$ yields to order n^{-1}

$$R_i^* = \sigma^{-1}\left\{\left(1 - \frac{\hat{v}_i}{2n\sigma^2}\right)R_i - \frac{\hat{m}_i}{n}\right\}.$$

Since $\hat{m}_i = m_i + \mathcal{O}_p(n^{-1/2})$ and $\hat{v}_i = v_i + \mathcal{O}_p(n^{-1/2})$, we can write R_i^* equivalently to order n^{-1} as

$$R_i^* = \sigma^{-1}\left\{R_i - n^{-1}\left(m_i + \frac{v_i R_i}{2\sigma^2}\right)\right\}, \tag{15}$$

which implies trivially that $E(R_i^*) = 0 + \mathcal{O}(n^{-3/2})$ and $\mathrm{Var}(R_i^*) = 1 + \mathcal{O}(n^{-3/2})$. Then, the adjusted residuals (14) have zero mean and unit variance to order n^{-1}.

Let $S_i = \{R_i - n^{-1}(m_i + \frac{v_i R_i}{2\sigma^2})\}$. Since $R_i = \mathcal{O}_p(1)$, the cumulative distribution function of S_i, $F_{S_i}(x)$ say, can be obtained from (15) to order n^{-1} following the approach developed by Cordeiro and Ferrari (1998a; 1998b, Section 2). We have

$$F_{S_i}(x) = F_{R_i}(x) + \frac{1}{n}\left(m_i + \frac{v_i x}{2\sigma^2}\right)f_{R_i}(x). \tag{16}$$

Differentiation of (16) with respect to x, and replacing $f_{R_i}(x)$ by its asymptotic expansion in (10), yields the density of S_i to order n^{-1}

$$f_{S_i}(x) = f_{\varepsilon_i}(x) - \frac{\mathrm{d}}{\mathrm{d}x}\{\theta_x^{(i)}f_{\varepsilon_i}(x)\} + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\{\phi_x^{(i)2}f_{\varepsilon_i}(x)\} + \frac{1}{n}\left\{\left(m_i + \frac{v_i x}{2\sigma^2}\right)\frac{\mathrm{d}f_{\varepsilon_i}(x)}{\mathrm{d}x} + \frac{v_i}{2\sigma^2}f_{\varepsilon_i}(x)\right\}. \tag{17}$$

The density function of R_i^* is $f_{R_i^*}(x) = \sigma f_{S_i}(\sigma x)$, where $f_{S_i}(\sigma x)$ comes from (17) with σx in place of x. The sum of the second and third terms of (17) are expressed as $\frac{\mathrm{d}}{\mathrm{d}x}\{\rho_i(x)f_{\varepsilon_i}(x)\}$. Since m_i/n, v_i/n, $\theta_x^{(i)}$ and $\phi_x^{(i)2}$ are all quantities of order $\mathcal{O}(n^{-1})$, the terms on the right hand side of (17), except $f_{\varepsilon_i}(x)$, are of this order and then the densities $f_{R_i^*}(x)$ and $f_{\varepsilon_i}(x)$ differ by terms of order $\mathcal{O}(n^{-1})$. However, we showed in Section 3, that the densities $f_{R_i'}(x)$ and $f_{\varepsilon_i}(x)$ are equal to this order. Thus, the distribution of the corrected residuals R_i', even in small samples, must be closer to the distribution of the true Pearson residuals than the distribution of the adjusted residuals R_i^*.

We now give a simple example of the expansion for the density $f_{R_i^*}(x)$ of the adjusted residuals R_i^* to order n^{-1} for the normal model with any link function. We have

$$f_{R_i^*}(x) = \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\left(1 + \frac{3\mu_i'^2 z_{ii}}{2} + \frac{v_i}{2n\sigma^2} - \frac{m_i}{\sigma n}x - \frac{\mu_i' B(\hat{\eta}_i)}{\sigma}x - \frac{\sigma\mu_i'' z_{ii}}{2}x - \frac{v_i}{2n\sigma^2}x^2 - \frac{3\mu_i'^2 z_{ii}}{2}x^2\right).$$

6. Simulation results

In this section some simulation results are presented to study the finite-sample distributions of the Pearson residual R_i, its corrected version R_i', its adjusted version R_i^* proposed by Cordeiro (2004) and the true Pearson residual. We use a gamma model with log link

$$\log\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

where the true parameters were taken as $\beta_0 = 1/2$, $\beta_1 = 1$, $\beta_2 = -1$ and $\phi = 4$. The explanatory variables x_1 and x_2 were generated from the uniform U(0, 1) distribution for $n = 20$ and their values were held constant throughout the

Table 4
Comparison of the mean and variance of Pearson, corrected, adjusted and true residuals.

i	Mean				Variance			
	R_i	R'_i	R^*_i	ε_i	R_i	R'_i	R^*_i	ε_i
1	0.013	0.006	0.011	0.004	0.234	0.255	1.059	0.257
2	−0.010	−0.006	0.007	0.001	0.183	0.232	1.112	0.255
3	0.002	−0.002	−0.004	−0.002	0.220	0.248	1.040	0.254
4	0.006	0.003	0.010	0.005	0.208	0.241	1.051	0.253
5	0.015	0.004	0.005	0.002	0.237	0.247	1.006	0.249
6	−0.003	−0.005	0.003	−0.001	0.188	0.229	1.043	0.245
7	−0.002	−0.006	−0.008	−0.005	0.201	0.237	1.038	0.244
8	−0.012	−0.009	0.001	−0.001	0.180	0.230	1.107	0.258
9	0.000	−0.001	0.008	0.000	0.201	0.244	1.087	0.253
10	−0.005	−0.010	−0.014	−0.010	0.207	0.235	0.999	0.236
11	−0.000	0.001	0.010	0.002	0.201	0.244	1.079	0.254
12	0.010	−0.001	−0.006	−0.000	0.243	0.252	1.022	0.259
13	−0.009	−0.012	−0.016	−0.009	0.199	0.230	1.002	0.239
14	−0.003	−0.001	0.019	0.005	0.176	0.227	1.116	0.248
15	0.012	0.005	0.010	0.004	0.221	0.243	1.017	0.252
16	−0.017	−0.017	−0.014	−0.007	0.174	0.225	1.105	0.249
17	0.004	−0.004	−0.009	−0.004	0.221	0.241	1.022	0.246
18	0.001	−0.004	−0.005	−0.004	0.214	0.240	1.020	0.246
19	0.000	−0.004	−0.006	−0.002	0.215	0.239	1.019	0.249
20	−0.003	−0.004	−0.003	−0.002	0.196	0.230	1.008	0.240

simulations. The total number of Monte Carlo replications was set at 10,000. All simulations were performed using the statistical software R.

In each of the 10,000 replications, we fitted the model and computed the MLEs $\hat{\beta}$ and $\hat{\mu}$, the Pearson residuals R_i, the function $\rho(\cdot)$ for this model and the corrected Pearson residuals R'_i. Further, we computed the expected values and variances from the expressions given by Cordeiro (2004) to obtain the adjusted Pearson residuals R^*_i and calculated the true Pearson residuals ε_i. Tables 4 and 5 give the sample mean and variance and the sample skewness and kurtosis of the residuals R_i, R'_i, R^*_i and ε_i, respectively, out of the 10,000 simulated values. Good agreement with the normal distribution happens when these figures are, on average, close to 0, 1, 0 and 3, respectively. The corrected residuals R'_i should agree with the true Pearson residuals rather than to the normal distribution.

The figures in Tables 4 and 5 show that the distribution of the corrected Pearson residuals is generally closer to the distribution of the true residuals than the distribution of the Pearson residuals. The correction $\rho(.)$ seems to be effective even when the sample size is small. The distribution of all residuals for the gamma model are positively skewed. All four cumulants of the corrected Pearson residuals R'_i are generally closer to the corresponding cumulants of the true Pearson residuals ε_i than those of the other residuals. The adjusted Pearson residuals R^*_i have cumulants much closer to the cumulants of the standard normal distribution as argued by Cordeiro (2004).

Table 6 gives the values of the one-sample Kolmogorov–Smirnov (K–S) and Anderson–Darling (A–D) distances (see, for instance, Anderson and Darling (1952) and Thode (2002, Section 5.1.4)) between the empirical distribution of the uncorrected and corrected residuals and the estimated distribution of the true residuals (a shifted gamma). The values of the K–S and A–D statistics measure the distances between the empirical distribution of each set of 10,000 uncorrected residuals R_i and corrected residuals R'_i, for $i = 1, \ldots, 20$, and the estimated distribution of the true residuals. Here, the estimated distribution is the shifted gamma distribution with dispersion parameter ϕ estimated by the sample mean of the estimates of the dispersion parameter at each step of the Monte Carlo experiment.

We are now interested in checking if the empirical distributions of the uncorrected R_i and corrected R'_i residuals have the same empirical distribution of the true residuals ε_i. Hence, we give in Table 7 both two-sample K–S and A–D distances between the empirical distribution of the uncorrected and corrected residuals and the empirical distribution of the true residuals.

The figures in Tables 6 and 7 indicate that the empirical distribution of the corrected residuals R'_i is closer to the distribution of the true residuals than the empirical distribution of the uncorrected residuals R_i, since the values of the K–S and A–D distances for the corrected residuals are substantially smaller than the corresponding distances for the uncorrected ones. This fact indicates that the corrected residuals represent an improvement over the uncorrected residuals when the model is well-specified.

We now provide an application of the corrected residuals in order to assess the adequacy of the gamma model. Under a well-specified model, we expect that the distribution of the corrected residuals will have approximately the distribution of the true residuals. However, even though it is common to compare the distribution of the Pearson residuals with the normal distribution, the standard normal approximation could not be adequate in small samples. Hence, we compare the empirical distribution of the corrected residuals with the distribution of the true residuals and the empirical distribution of the uncorrected residuals with the normal distribution. We use QQ plots of the sample quantiles of the corrected residuals versus the theoretical quantiles from the estimated distribution of the true residuals and of the sample quantiles of the

Table 5
Comparison of the skewness and kurtosis of Pearson, corrected, adjusted and true residuals.

i	Skewness				Kurtosis			
	R_i	R'_i	R^*_i	ε_i	R_i	R'_i	R^*_i	ε_i
1	0.837	0.943	0.626	1.005	3.798	4.105	2.967	4.468
2	0.586	0.822	0.494	0.986	3.205	3.780	2.805	4.399
3	0.824	0.973	0.605	1.080	3.898	4.350	3.020	4.859
4	0.703	0.863	0.550	0.979	3.417	3.825	2.882	4.395
5	0.876	0.942	0.626	0.964	4.040	4.232	3.012	4.275
6	0.628	0.829	0.523	0.987	3.278	3.772	2.823	4.387
7	0.715	0.901	0.548	0.960	3.548	4.052	2.920	4.317
8	0.611	0.864	0.500	1.068	3.318	3.984	2.865	4.813
9	0.711	0.923	0.557	1.017	3.561	4.162	2.911	4.667
10	0.809	0.965	0.628	1.018	3.904	4.387	3.061	4.811
11	0.727	0.936	0.556	1.018	3.590	4.176	2.920	4.532
12	0.939	1.001	0.659	1.077	4.361	4.560	3.076	4.929
13	0.746	0.907	0.603	0.938	3.607	4.052	3.006	4.106
14	0.553	0.801	0.474	0.939	3.150	3.709	2.820	4.254
15	0.808	0.928	0.606	1.048	3.813	4.154	3.033	4.737
16	0.593	0.851	0.506	1.006	3.246	3.833	2.859	4.510
17	0.793	0.910	0.606	0.958	3.727	4.058	2.994	4.202
18	0.783	0.923	0.610	0.992	3.686	4.078	2.977	4.411
19	0.776	0.904	0.603	0.963	3.687	4.060	2.993	4.292
20	0.715	0.888	0.569	0.963	3.532	4.004	2.960	4.346

Table 6
One-sample K–S and A–D statistics on Pearson and corrected residuals.

i	K–S stat. for R_i	A–D stat. for R_i	K–S stat. for R'_i	A–D stat. for R'_i
All	0.0232	300.42	0.0036	6.7641
1	0.0230	7.5944	0.0103	1.7875
2	0.0317	30.7031	0.0074	1.2504
3	0.0208	7.9810	0.0077	1.0710
4	0.0287	17.5666	0.0100	1.2283
5	0.0216	8.9498	0.0098	1.1761
6	0.0307	28.3464	0.0074	0.8353
7	0.0273	17.8230	0.0088	1.0719
8	0.0311	34.7206	0.0109	1.7666
9	0.0277	19.3796	0.0081	0.9986
10	0.0244	12.6919	0.0123	1.9530
11	0.0306	19.5087	0.0089	0.6709
12	0.0167	3.6271	0.0106	2.0631
13	0.0208	15.1071	0.0107	2.2356
14	0.0401	49.8411	0.0117	1.6905
15	0.0277	16.1354	0.0150	1.8709
16	0.0360	43.4746	0.0155	2.5022
17	0.0186	7.3023	0.0082	1.0216
18	0.0235	9.8595	0.0068	0.6481
19	0.0172	7.6480	0.0085	0.8816
20	0.0282	21.7356	0.0072	0.7146

uncorrected residuals versus the theoretical quantiles from the normal distribution with zero mean and variance $1/\hat{\phi}$. If the distribution of the corrected residuals is close to the distribution of the true residuals, the plot should form a straight line. Thus, we expect that the QQ plot of the corrected residuals versus the estimated distribution of the true residuals should be closer to the diagonal line than the QQ plot of the uncorrected residuals against the normal distribution mentioned above. As a further study, we also provide the QQ plot of the adjusted residuals against the theoretical quantiles of the standard normal distribution.

Fig. 1 gives two QQ plots, one for the vector of the 10,000 ordered uncorrected residuals and other for the vector of the 10,000 ordered corrected residuals. These plots show that even for a well-specified model, the uncorrected residuals display large deviations from the diagonal line when compared to the corrected residuals. In fact, the plotted points for the corrected residuals appear to cluster around the straight line drawn through them, visually supportive evidence that these residuals come from a distribution which can be adequately approximated by the estimated shifted gamma distribution. The plot for the adjusted residuals given in Fig. 2 it is clearly an improvement on the plot of the uncorrected residuals, but it shows substantive deviations from the diagonal line when compared to the corrected residuals. Hence, the corrected residuals behave very well and lead to the right conclusion, i.e., that the model is well-specified. We then recommend the corrected residuals to build up QQ plots.

Table 7
Two-sample K–S and A–D statistics on Pearson and corrected residuals.

i	K–S stat. for R_i	A–D stat. for R_i	K–S stat. for R_i'	A–D stat. for R_i'
All	0.0283	444.193	0.0041	9.5703
1	0.0246	10.5690	0.0086	0.5191
2	0.0356	44.3716	0.0125	1.1887
3	0.0273	15.8880	0.0100	0.8209
4	0.0331	22.3655	0.0079	0.9083
5	0.0227	9.41018	0.0070	0.4727
6	0.0355	33.9614	0.0094	1.5002
7	0.0348	23.8043	0.0077	0.5297
8	0.0394	54.3072	0.0126	1.3377
9	0.0325	27.3448	0.0071	0.5715
10	0.0270	14.1363	0.0065	0.2252
11	0.0336	28.1285	0.0100	0.8706
12	0.0218	9.17441	0.0102	0.4995
13	0.0342	22.1684	0.0112	1.2824
14	0.0426	56.0126	0.0132	2.5277
15	0.0281	15.7758	0.0121	1.1146
16	0.0444	57.0837	0.0109	1.9487
17	0.0255	14.0805	0.0081	0.5073
18	0.0310	16.4867	0.0089	0.4553
19	0.0282	16.7720	0.0094	1.0549
20	0.0303	22.7455	0.0087	0.8158

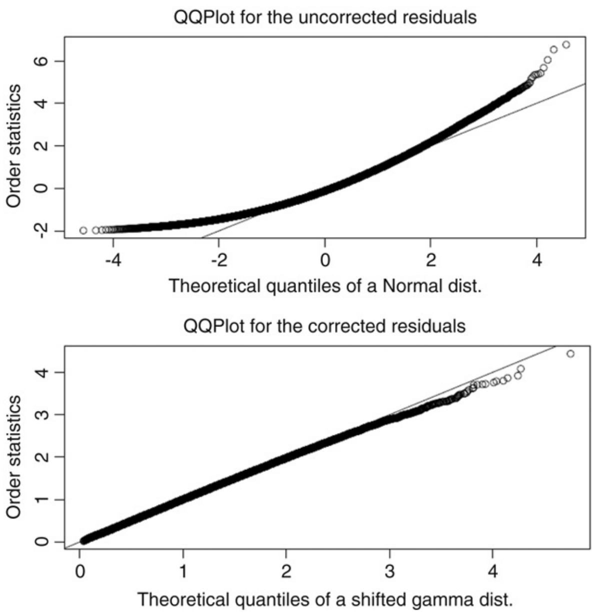

Fig. 1. QQ plots for the Pearson and corrected residuals.

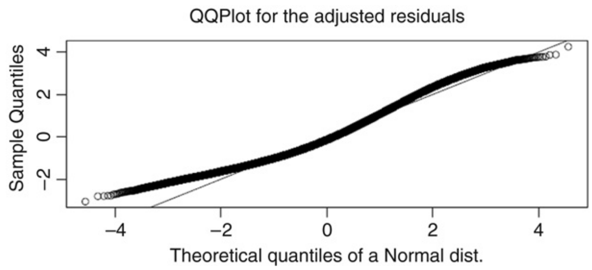

Fig. 2. QQ plot for the adjusted residuals.

Table 8
Times until loss of velocity for five types of turbines.

Type of turbine				
Type I	Type II	Type III	Type IV	Type V
3.03	3.19	3.46	5.88	6.43
5.53	4.26	5.22	6.74	9.97
5.60	4.47	5.69	6.90	10.39
9.30	4.53	6.54	6.98	13.55
9.92	4.67	9.16	7.21	14.45
12.51	4.69	9.40	8.14	14.72
12.95	5.78	10.19	8.59	16.81
15.21	6.79	10.71	9.80	18.39
16.04	9.37	12.58	12.28	20.84
16.84	12.75	13.41	25.46	21.51

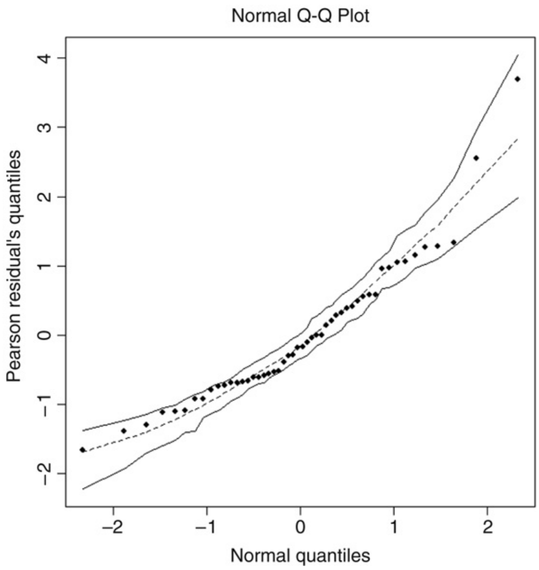

Fig. 3. QQ plot for the unadjusted Pearson residuals in the complete model.

7. Application to real data

We now show the usefulness of the corrected residuals with an application to a real data set. The data given in Table 8 is taken from Lawless (1982, p. 201) and consist of the times to evaluate the performance of five types of high-speed turbines for plane engines. Ten engines of each type were considered in the analysis and the times (in million of cycles units) until the loss of velocity are recorded.

Let T_{ij} be the time until the loss of velocity of the jth engine for the ith type of turbine, $i = 1, \dots, 5$ and $j = 1, \dots, 10$. We assume that T_{ij} has a gamma distribution with mean μ_{ij} and dispersion parameter ϕ^{-1}. To compare the five types of turbines we consider the one-way classification model

$$\mu_{ij} = \mu + \beta_i,$$

where $\beta_1 = 0$.

The deviance of the fitted model is 42.56 on 45 degrees of freedom yielding an adequate global fit. The estimates of the parameters with standard errors in parentheses are: $\hat{\mu} = 10.69$ (1.54), $\hat{\beta}_2 = -4.643$ (1.773), $\hat{\beta}_3 = -2.057$ (1.983), $\hat{\beta}_4 = -0.895$ (2.093) and $\hat{\beta}_5 = 4.013$ (2.624). The MLE of ϕ is $\hat{\phi} = 4.803$.

We use the Pearson residuals to check for discrepancies between the model and the data. Fig. 3 exhibits the QQ plot for the quantiles of the unadjusted Pearson residuals against the quantiles of a normal distribution with the corresponding numerical envelope (see, for instance, Atkinson (1985)). The numerical envelope is used to circumvent the fact that the unadjusted Pearson residuals do not have a normal distribution. Fig. 3 shows that there are no discrepancies between the model and the data, thus indicating a good fit.

Fig. 4 exhibits the QQ plot for the quantiles of the corrected Pearson residuals against the quantiles of the estimated shifted gamma distribution. This plot indicates that there are two outliers, namely the observations 20 and 40, but except for these two observations, the complete model seems adequate to fit the data. To verify this fact, Fig. 5 gives the plot of

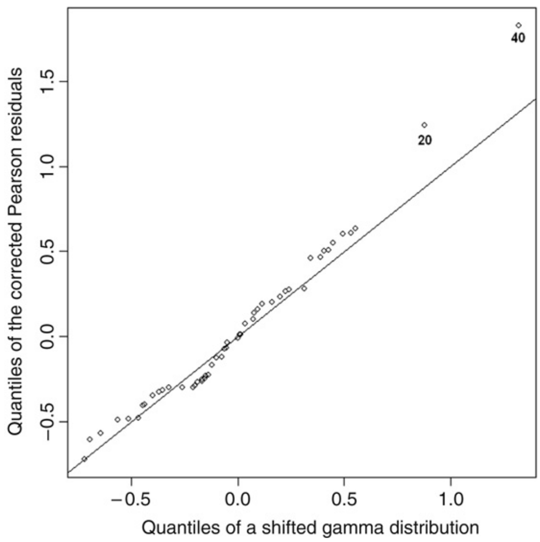

Fig. 4. QQ plot for the corrected Pearson residuals in the complete model.

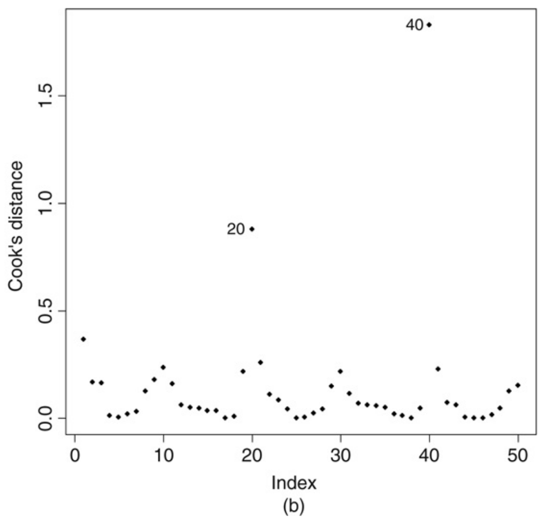

Fig. 5. Plot of the Cook's distance.

Cook's distance versus the index of the observation to detect possible influent observations, suggesting that the observations 20 and 40 are influent as detected before using the corrected residuals.

The analysis discussed before shows that the corrected residuals can be useful to provide diagnostic tools. Moreover, since we are dealing with the true distribution of the residuals to order $\mathcal{O}(n^{-1})$, the numerical envelope could not be necessary, since we can obtain much more information from the QQ plot.

We go further to study the inferential aspects on the estimates and an appropriate test of interest could be $\mathcal{H}_0 : \beta_3 = \beta_4 = 0$ against $\mathcal{H}_1 : \beta_3 \neq 0$ or $\beta_4 \neq 0$. For testing $\mathcal{H}_0$, we obtain a p-value 0.562 which indicates that this hypothesis should not be rejected.

If we eliminate both observations 20 and 40, the conclusion of the above test will not change. Hence, we estimate a new model with $\beta_1 = \beta_3 = \beta_4 = 0$. As suggested by both procedures, i.e. the corrected residuals and Cook's distances, we estimate this new model without the observations 20 and 40. For the fitted reduced model without these observations, we obtain $\hat{\mu} = 9.1659\,(0.6412)$, $\hat{\beta}_2 = -3.8603\,(0.9247)$ and $\hat{\beta}_5 = 5.5401\,(1.8657)$ and the MLE of ϕ is now $\hat{\phi} = 7.045474$.

Fig. 6 exhibits the QQ plot for the quantiles of the unadjusted Pearson residuals against the quantiles of a normal distribution with the corresponding numerical envelope. As before, this plot suggests that the reduced model is well-specified. Fig. 7 gives the QQ plot for the quantiles of the corrected Pearson residuals against the quantiles of a shifted gamma distribution. We also conclude that the reduced model is satisfactory and there are no influent observations.

Finally, we conclude from this application that the unadjusted Pearson residuals seem to be too optimistic and would lead a practitioner to stop the investigation in an early stage of the model selection, since the influent observations could deserve

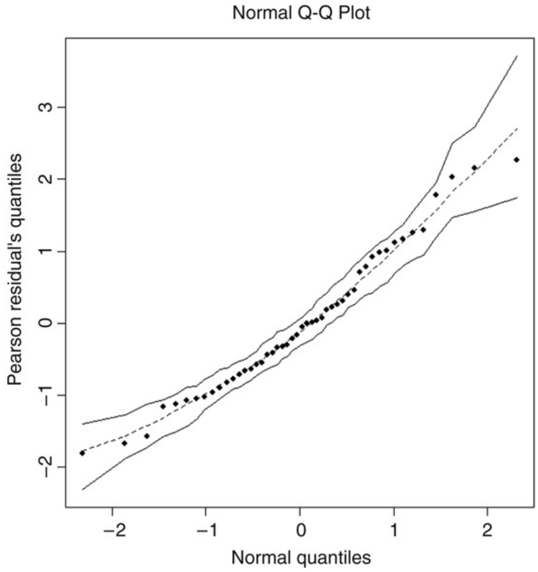

Fig. 6. QQ plot for the unadjusted Pearson residuals in the reduced model.

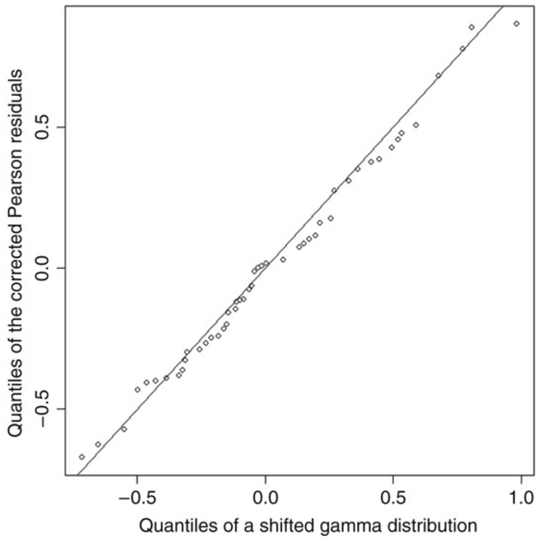

Fig. 7. QQ plot for the corrected Pearson residuals in the reduced model.

a further investigation. Nevertheless, the corrected residuals in this case determine precisely the influent observations as they were detected by the Cook's distance plot.

8. Conclusion

In regression models, Pearson residuals are either compared with quantiles of the standard normal distribution or analyzed with the aid of residual plots with simulated envelopes. However, the normal approximation is not adequate in small samples. For the first time, we obtain the density of Pearson residuals in continuous generalized linear models corrected to order n^{-1}, where n is the sample size, and define corrected residuals for these models which have the same distribution of the true residuals to this order of approximation. The setup is similar to the paper by Loynes (1969), but applied in a wide context of regression models, and Pierce and Schafer (1986). The article can also be considered a sequel to Cordeiro (2004). We provide applications to some common models. The performance of the uncorrected Pearson, corrected and adjusted residuals proposed by Cordeiro (2004) are compared in a simulation study. Under a well-specified gamma model, we show by simulation that the QQ plot of the corrected residuals versus the estimated distribution of the true residuals is much closer to the straight line than the QQ plot of the uncorrected residuals against the normal distribution.

Moreover, in an example applied to a real data set, we demonstrate that the corrected residuals are better than the uncorrected Pearson residuals to identify discrepancies between the fitted model and the data.

Acknowledgment

We are very grateful to two referees and the Editor for helpful comments that considerably improved the paper. We gratefully acknowledge financial support from CNPq.

Appendix

Suppose we write the residual R in terms of the true residual ε as $R = \varepsilon + \delta$, where ε and δ are absolutely continuous random variables with respect to the Lebesgue measure and δ is of order $\mathcal{O}_p(n^{-1})$. Our goal is to define a corrected residual R' having the same density of ε to order n^{-1}. Initially, we have

$$E(\mathrm{e}^{\mathrm{i}sR}) = E\{\mathrm{e}^{\mathrm{i}s\varepsilon}E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon)\} \quad \text{and} \quad \left.\frac{\partial^k}{\partial s^k}E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon)\right|_{s=0} = \mathrm{i}^k E(\delta^k \mid \varepsilon).$$

Expanding $E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon)$ in a Taylor series around $s = 0$ gives

$$E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon) = 1 + (\mathrm{i}s)E(\delta \mid \varepsilon) + \frac{(\mathrm{i}s)^2}{2}E(\delta^2 \mid \varepsilon) + \cdots.$$

Let $\theta_x = E(\delta \mid \varepsilon = x)$ and $\phi_x^2 = \mathrm{Var}(\delta \mid \varepsilon = x)$. Thus,

$$E\{\mathrm{e}^{\mathrm{i}s\varepsilon}E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon)\} = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}sx}\left\{1 + (is)\theta_x + \frac{(\mathrm{i}s)^2}{2}(\phi_x^2 + \theta_x^2) + \cdots\right\} f_\varepsilon(x)\mathrm{d}x, \tag{18}$$

where $f_\varepsilon(\cdot)$ is the density function of ε. Using Cox and Snell's (1968) formulae (25) and (26) with $\varepsilon = 0$, we conclude that $E(\delta)$ and $\mathrm{Var}(\delta)$ (and thus $E(\delta^2)$) are of order $\mathcal{O}(n^{-1})$ and, in the same way, that the higher moments of δ are of order $o(n^{-1})$. In a similar manner, we can show that $E(\delta \mid \varepsilon = x)$ and $\mathrm{Var}(\delta \mid \varepsilon = x)$ are of order $\mathcal{O}(n^{-1})$, and also that the higher-order conditional moments are of order $o(n^{-1})$. Hence, we can rewrite Eq. (18) as

$$E\{\mathrm{e}^{\mathrm{i}s\varepsilon}E(\mathrm{e}^{\mathrm{i}s\delta} \mid \varepsilon)\} = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}sx}\left\{1 + (\mathrm{i}s)\theta_x + \frac{(\mathrm{i}s)^2}{2}\phi_x^2\right\} f_\varepsilon(x)\mathrm{d}x + o(n^{-1}). \tag{19}$$

We can express the integral on the right side of (19) as a sum of three integrals. Then, integration by parts, one time for the integral containing θ_x and two times for the integral containing ϕ_x^2, gives

$$E(\mathrm{e}^{\mathrm{i}sR}) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}sx}\left[f_\varepsilon(x) - \frac{\mathrm{d}}{\mathrm{d}x}\{f_\varepsilon(x)\theta_x\} + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\{f_\varepsilon(x)\phi_x^2\}\right]\mathrm{d}x + o(n^{-1}). \tag{20}$$

The uniqueness theorem for characteristic functions yields the density of R to order n^{-1}

$$f_R(x) = f_\varepsilon(x) - \frac{\mathrm{d}}{\mathrm{d}x}\{f_\varepsilon(x)\theta_x\} + \frac{1}{2}\frac{d^2}{\mathrm{d}x^2}\{f_\varepsilon(x)\phi_x\} + o(n^{-1}). \tag{21}$$

Equation (21) is identical to formula (5) in Loynes (1969).

Further, we define corrected residuals of the form $R' = R + \rho(R)$, where $\rho(\cdot)$ is a function of order $\mathcal{O}(n^{-1})$ used to recover the distribution of ε. We may proceed as above, noting that $E\{\rho(R) \mid R = x\} = \rho(x)$, to obtain the density of R' to order n^{-1}

$$f_{R'}(x) = f_R(x) - \frac{\mathrm{d}}{\mathrm{d}x}\{\rho(x)f_R(x)\}.$$

Since the quantities $\rho(x)$, θ_x and ϕ_x^2 are all of order $\mathcal{O}(n^{-1})$, we have to this order that $\frac{\mathrm{d}}{\mathrm{d}x}\{\rho(x)f_R(x)\} = \frac{\mathrm{d}}{\mathrm{d}x}\{\rho(x)f_\varepsilon(x)\}$. Hence, the density functions of R and ε will be the same to order n^{-1} if

$$\frac{\mathrm{d}}{\mathrm{d}x}\{\rho(x)f_\varepsilon(x)\} = -\frac{\mathrm{d}}{\mathrm{d}x}\{f_\varepsilon(x)\theta_x\} + \frac{1}{2}\frac{d^2}{\mathrm{d}x^2}\{f_\varepsilon(x)\phi_x\}.$$

Integration gives

$$\rho(x) = -\theta_x + \frac{1}{2f_\varepsilon(x)}\frac{\mathrm{d}}{\mathrm{d}x}\{f_\varepsilon(x)\phi_x\}. \tag{22}$$

Eq. (22) is precisely equation (6) given in Loynes (1969) and it is clear from the proof that the support of ε does not need to be the entire line. We can have proper intervals as support. We should note that the assumptions needed can be made weaker if we require that an expansion of the Taylor polynomial of order two with a remainder term (for instance, Lagrange remainder) can be taken instead of the complete series.

We could also prove Loynes (1969) results by using the equivalence of (3c) and (4c), together with (5) and (6) of Cox and Reid (1987) and appropriate regularity conditions. The idea of this approach is as follows: consider in equation (3c) of Cox and Reid (1987) $X_0 = \varepsilon, X_1 = n^{1/2}\delta$ and $X_2 = 0$. This means that we are writing Y_n as $Y_n = \varepsilon + \delta + \mathcal{O}_p(n^{-3/2})$, where ε is $\mathcal{O}_p(1)$ and δ is $\mathcal{O}_p(n^{-1})$. Then, by (4c), (5) and (6) of Cox and Reid (1987), we can write the cumulative distribution function (cdf) of Y_n as

$$G_n(y) = F_0(y) - E(\delta \mid \varepsilon = y)f_0(y) + \frac{1}{2}\frac{\partial}{\partial y}\{E(\delta^2 \mid \varepsilon = x)f_0(y)\} + \mathcal{O}(n^{-3/2}),$$

where $F_0(\cdot)$ and $f_0(\cdot)$ are the cdf and density function of ε, respectively. This equation implies expression (21). We can also obtain the expansion for $R + \rho(R)$ from the equivalence of (3c) and (4c) of Cox and Reid (1987) by setting $X_0 = R, X_1 = 0$ and $X_2 = \rho(R)$. The rest of the proof is identical to the one gave above. Note also, that this proof does not require ε to be supported in the entire line since this is not needed in the regularity conditions.

References

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann. Math. Statist. 23, 193–212.
Atkinson, A.C., 1985. Plots, Transformations, and Regression. Oxford University Press, London.
Cordeiro, G.M., 2004. On Pearson's residuals in generalized linear models. Statist. Probab. Lett. 66, 213–219.
Cordeiro, G.M., Ferrari, S.L.P., 1998a. A note on Bartlett-type correction for the first few moments of test statistics. J. Statist. Plann. Inference 71, 261–269.
Cordeiro, G.M., Ferrari, S.L.P., 1998b. Generalized Bartlett corrections. Comm. Statist. Theory Methods 27, 509–527.
Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. J. Roy. Statist. Soc. B 53, 629–643.
Cordeiro, G.M., Paula, G.A., 1989. Improved likelihood ratio statistic for exponential family nonlinear models. Biometrika 76, 93–100.
Cox, D.R., Reid, N., 1987. Approximations to noncentral distributions. Canad. J. Statist. 15, 105–114.
Cox, D.R., Snell, E.J., 1968. A general definition of residuals (with discussion). J. Roy. Statist. Soc. B 30, 248–275.
Lawless, J.F., 1982. Statistical Models and Methods for Lifetime Data. John Wiley, New York.
Loynes, R.M., 1969. On Cox and Snell's general definition of residuals. J. Roy. Statist. Soc. B 31, 103–106.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, London.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. J. Roy. Statist. Soc. A 135, 370–384.
Pierce, D.A., Schafer, D.W., 1986. Residuals in generalized linear models. J. Amer. Statist. Assoc. 81, 977–986.
Pregibon, D., 1981. Logistic regression diagnostics. Ann. Statist. 9, 705–724.
Simas, A.B., Cordeiro, G.M., 2009. Adjusted Pearson residuals in exponential family nonlinear models. J. Statist. Comput. Simul. 79, 411–425.
Thode Jr., H.C., 2002. Testing for Normality. Marcel Dekker, New York.
Wei, B.-C., 1998. Exponential Family Nonlinear Models. Springer, Singapore.
Williams, D.A., 1984. Residuals in generalized linear models. In: Proc. 12th International Biometrics Conference, Tokyo, pp. 59–68.
Williams, D.A., 1987. Generalized linear model diagnostics using the deviance and single case deletions. Appl. Statist. 36, 181–191.