Testing the Poisson Regression Model

J.C.W. Rayner and Paul Rippon

The University of Newcastle, Callaghan, NSW, 2308, AUSTRALIA
John.Rayner@newcastle.edu.au and Paul.Rippon@newcastle.edu.au

Abstract

A smooth test of the Poisson assumption in the Poisson regression generalised linear model is derived.

Keywords: Generalised linear models, goodness of fit, leverage matrix, smooth tests

1. Introduction

The smooth tests in Rayner et al. (2009) assume observations are independent and identically distributed. For example, section 2 here recalls the smooth test for the Poisson distribution. In section 3 we describe the Poisson regression model where observations are *not* independent and identically distributed. The smooth test for the Poisson assumption is outlined; details are given in the Appendix. Similarities and differences in the two tests are discussed in section 4.

We note that the smooth test derived here can be extended to generalised linear models in general. See Pena and Slate (2006) who consider smooth testing in making a global assessment of the assumptions for the linear model.

2. Testing the Poisson Assumption for independent and identically distributed observations

In deriving their smooth tests that data are consistent with a specified probability density function, Rayner et al. (2009) assume a random sample $Y_1, \ldots, Y_n$ and nest the probability density function in a smooth alternative. When the null distribution is Poisson with positive mean μ, the probability density function is, for $y = 0, 1, \ldots$, $f(y;\ \mu) = \exp(-\mu)\mu^y/y!$. A smooth alternative of order k is

$$C(\boldsymbol{\theta}, \mu)\exp\left\{\sum_{i=1}^{k} \theta_i h_i(y;\mu)\right\} f(y;\mu)$$

in which $\boldsymbol{\theta} = (\theta_i)$ and $C(\boldsymbol{\theta},\ \mu)$ is a normalising constant that is assumed to exist and satisfies

$$C(\boldsymbol{\theta}, \mu)\sum_{y=0}^{\infty} \exp\left\{\sum_{i=1}^{k} \theta_i h(y;\mu)\right\} f(y;\mu) = 1. \quad (1)$$

Here $\{h_i(y;\ \mu)\}$ is the set of Poisson-Charlier polynomials; see Rayner et al. (2009, p.155). The first three polynomials are $h_0(y;\ \mu) = 1$, $h_1(y;\ \mu) = (y - \mu)/\sqrt{\mu}$ and $h_2(y;\ \mu) = \{(y - \mu)^2 - y\}/\sqrt{(2\mu^2)}$. We seek to assess the Poisson assumption by testing H: $\boldsymbol{\theta} = 0$ against K: $\boldsymbol{\theta} \neq 0$ using the score test for this model.

With the model as given, the score test statistic is found to have singular covariance matrix. One solution is to remove $\theta_1 h_1(y;\ \mu)$ from the smooth alternative. It seems that θ_1 and μ are fulfilling the same role in the model, and if both are included one is redundant. If the new order k smooth alternative includes θ_2 to θ_{k+1} then the order k score test statistic is

$$\sum_{i=2}^{k+1} V_i^2 \text{ where } V_i = \sum_{j=1}^{n} h_i\left(Y_j;\hat{\mu}\right)/\sqrt{n}$$

in which $\hat{\mu}$ is the maximum likelihood estimator of μ.

3. A Smooth Test For Poisson Regression

According to Dobson (2002, p.152) the Poisson regression model is defined when $Y_1, \ldots, Y_n$ are independent Poisson (μ_j) random variables with $E[Y_j] = \mu_j = \exp(\mathbf{x}_j^T \boldsymbol{\beta})$ in which $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})^T$, $j = 1, \ldots, n$, are $p \times 1$ vectors of constants and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of nuisance parameters. Dobson (2002, p.152) includes an offset term, but here this term will be absorbed into μ_j. To confirm the model a smooth alternative likelihood L is constructed:

$$L = \prod_{j=1}^{n} C(\boldsymbol{\theta}, \mu_j) \exp\left\{\sum_{i=1}^{k} \theta_i h_i(y_j; \mu_j)\right\} f(y_j; \mu_j)$$

$$= \left\{\prod_{j=1}^{n} C(\boldsymbol{\theta}, \mu_j)\right\} \exp\left\{\sum_{i=1}^{k} \theta_i \sum_{j=1}^{n} h_i(y_j; \mu_j)\right\} \times$$

$$\exp\left\{-\sum_{j=1}^{n} \mu_j\right\}\left\{\prod_{j=1}^{n} \mu_j^{y_j}\right\} / \left\{\prod_{j=1}^{n} y_j!\right\}$$

in which, as before, $\{h_i(y; \mu)\}$ is the set of Poisson-Charlier polynomials and $C(\boldsymbol{\theta}, \mu)$ is a normalising constant that is assumed to exist. When $\boldsymbol{\theta} = 0$, L is the likelihood of the Poisson regression model.

To derive the score test statistic, derivatives to second order of the logarithm of the likelihood are required, and these in turn require derivatives to second order of $C(\boldsymbol{\theta}, \mu_j)$. These can be found by differentiating equation (1). The details are given in the Appendix. Define $V_u = \sum_{j=1}^{n} h_u(y_j; \mu_j)/\sqrt{n}$. Henceforth parameters estimated under the null hypothesis have a zero subscript. Now under H: $\boldsymbol{\theta} = 0$,

$$\frac{\partial \log L}{\partial \theta_u} = V_u\sqrt{n} \text{ for } u = 1, \ldots, k, \text{ and}$$

$$\frac{\partial \log L}{\partial \beta_s} = -\sum_{j=1}^{n} \{y_j - \mu_{0j}\} x_{js} \quad \text{for } s = 1, \ldots, p.$$

The nuisance parameters are found by solving $\partial \log L / \partial \beta = 0$. Dobson (2002, section 4.3, pp.62) described finding the ML estimates using the *method of scoring*. From Dobson (2002, 4.19) the information matrix is found to be $\left(\sum_{j=1}^{n} \mu_{0j} x_{js} x_{jt}\right) = \mathbf{X}^T\mathbf{D}\mathbf{X}$ where $\mathbf{X} = (x_{js})$. This is consistent with our subsequent results. Having estimated $\boldsymbol{\beta}$ in this way $\hat{\boldsymbol{\mu}}_0$ can be determined from $\log \mu_j = x_j^T \hat{\beta}_0$ for all j. These estimators are needed, for example, to calculate the V_u.

We now write $\mathbf{D} = \mathrm{diag}(\mu_{01}, \ldots, \mu_{0n})$, $\mathbf{I}_k$ is the $k \times k$ identity matrix, $\mathbf{1}_n$ for the $n \times 1$ vector of 1s and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$. The partitioned information matrix has blocks $\mathbf{I}_{\theta\theta}$, $\mathbf{I}_{\theta\beta}$, $\mathbf{I}_{\beta\theta} = \mathbf{I}_{\theta\beta}^T$ and $\mathbf{I}_{\beta\beta}$, given by

$$\mathbf{I}_{\theta\theta} = n\,\mathbf{I}_k,\ \mathbf{I}_{\theta\beta} = \begin{pmatrix} \mathbf{1}_n^T \mathbf{D}^{1/2}\mathbf{X} \\ \mathbf{0} \end{pmatrix} \text{ and } \mathbf{I}_{\beta\beta} = \mathbf{X}^T\mathbf{D}\mathbf{X}.$$

The asymptotic covariance matrix is $\boldsymbol{\Sigma} = \mathbf{I}_{\theta\theta} - \mathbf{I}_{\theta\beta}\mathbf{I}_{\beta\beta}^{-1}\mathbf{I}_{\beta\theta}$. In the resulting $\mathbf{I}_{\theta\beta}\mathbf{I}_{\beta\beta}^{-1}\mathbf{I}_{\beta\theta}$, only the (1, 1)th term will be non-zero. It follows that

$$(\boldsymbol{\Sigma})_{11} = \mathbf{1}_n^T \{\mathbf{I}_n - \mathbf{D}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{D}\mathbf{X})^{-1}\mathbf{X}^T \mathbf{D}^{1/2}\}\mathbf{1}_n = n\sigma^2 \text{ say.}$$

All off-diagonal elements of $\boldsymbol{\Sigma}$ are zero, and all remaining diagonal elements are n. The score test statistic is thus

$$V_1^2/\sigma^2 + V_2^2 + \ldots + V_k^2$$

in which all parameters are estimated by maximum likelihood. The matrix $\mathbf{D}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{D}\mathbf{X})^{-1}\mathbf{X}^T \mathbf{D}^{1/2}$ that appears in $(\boldsymbol{\Sigma})_{11}$ is the familiar leverage matrix for this model; see Dean and Lawless (1989) and the references therein. Note that Dean and Lawless (1989) discuss tests for extra-Poisson variation including one based on V_2^2.

Under the null hypothesis the smooth test statistic has the χ_k^2 distribution with k degrees of freedom. This is the number of parameters in the full model, k ($\theta_1, \ldots, \theta_k$) + p ($\beta_1, \ldots, \beta_p$), minus the number of parameters in the null model, p ($\beta_1, \ldots, \beta_p$). At first glance this seems a little surprising: no matter how many parameters β_j are used to specify the structure of the model, the degrees of freedom of the test that assesses the model remain the same.

4. Discussion

The essential difference between the smooth tests for the independent and identically distributed and Poisson regression models is that although both include a sum of squares of components from the second on, for the Poisson regression model a term involving V_1^2 is included, while the test for the independent and identically distributed model does not.

Rayner et al. (2009, section 9.2) show that for the independent and identically distributed model in testing for a distribution in an exponential family a smooth test statistic that is a sum of squares can readily be constructed. However in the Poisson regression model it is not possible to express the likelihood in the form of an exponential family. Nevertheless in an intuitive sense the V_2^2 + ... + V_k^2 part of the test statistic in Poisson regression is a carryover from the independent and identically distributed assessment.

Outside of exponential families the score test statistic is very occasionally a weighted sum of squares of components, as for the Laplace and logistic distributions (see Rayner et al, 2009, sections 11.2 and 11.3). It is more usually a quadratic form that allows no simplification. So here the weighted first component is not dissimilar to what happens for the Laplace and logistic distributions.

In the independent and identically distributed model there is only one location parameter. This is estimated by maximum likelihood, which in exponential families is equivalent to method of moments estimation, and for the Poisson model that means V_1 is identically zero. Since V_1 is degenerate its variance and covariances with the other V_r are all zero. This is why the asymptotic covariance matrix is singular and it doesn't make sense to include a θ_1 term in the smooth model.

Now in the Poisson regression case there are p βs to be estimated. If $p = 1$ and this is the offset term, then the model reduces to the independent and identically distributed model. With $p \geq 2$ the likelihood equations neither include nor imply $V_1 \equiv 0$. In fact the V_1 term may be thought of as assessing the ability of the link $\mu_j = \exp(\mathbf{x}_j^T\boldsymbol{\beta})$ to model location.

References

DEAN, C. and LAWLESS, J.F. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, 84, 467-472.

DOBSON, Annette J. (2002). *An Introduction to Generalised Linear Models*. Boca Raton: Chapman & Hall/CRC.

RAYNER, J.C.W., THAS, O. and BEST, D.J. (2009). *Smooth Tests of Goodness of Fit: Using R* (2nd ed.). Singapore: Wiley.

PENA, E.A. and SLATE, E.H. (2006). Global validation of linear model assumptions. *Journal of the American Statistical Association*, 101, 341-353.

RIPPON, Paul and RAYNER, J.C.W. (2009). Testing the distributional assumptions in Poisson regression. Companion paper.

Appendix

Since the smooth alternative distribution has a proper probability (mass) function

$$C(\boldsymbol{\theta},\mu)\sum_{y=0}^{\infty}\exp\left\{\sum_{i=1}^{k}\theta_i h(y;\mu)\right\}f(y;\mu) = 1. \quad (2)$$

Differentiating both sides of (2) with respect to θ_u and β_s gives, ultimately,

$$\frac{\partial \log C}{\partial \theta_u} = -\,\mathrm{E}[h_u(Y;\mu)],\ u = 1, \dots, k, \quad (3)$$

and

$$\frac{\partial \log C}{\partial \beta_s} = -\sum_{i=1}^{k}\theta_i \mathrm{E}\left[\frac{\partial h_i(Y;\mu_j)}{\partial \beta_s}\right] - \mathrm{E}\left[\frac{\partial \log f}{\partial \beta_s}\right],$$
$$s = 1, \dots, p. \quad (4)$$

Henceforth we write $C_j = C(\theta, \mu_j)$ for $j = 1, \dots, n$ and note that $\partial\mu_j/\partial\beta_s = \mu_j\, x_{js}$. Differentiating the logarithm of the likelihood gives, for $u = 1, \dots, k$ and $s = 1, \dots, p$

$$\frac{\partial \log L}{\partial \theta_u} = \sum_{j=1}^{n}\frac{\partial \log C_j}{\partial \theta_u} + \sum_{j=1}^{n} h_u(y_j;\mu_j), \text{ and}$$

$$\frac{\partial \log L}{\partial \beta_s} = \sum_{j=1}^{n}\frac{\partial \log C_j}{\partial \beta_s}$$
$$+ \sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}\frac{\partial h_i(y_j;\mu_j)}{\partial \beta_s} + \sum_{j=1}^{n}(y_j - \mu_j)x_{js}.$$

On using (3) and (4) we have

$$\frac{\partial \log L}{\partial \theta_u} = \sum_{j=1}^{n}\{h_u(y_j;\mu_j) - \mathrm{E}[h_u(Y;\mu_j)]\} \text{ and}$$

$$\frac{\partial \log L}{\partial \beta_s} = \sum_{j=1}^{n}\{y_j - \mu_j\}x_{js} - n\mathrm{E}\left[\frac{\partial \log f}{\partial \beta_s}\right]$$

$$\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}\left\{\frac{\partial h_i(y_j;\mu_j)}{\partial \beta_s} - \mathrm{E}\left[\frac{\partial h_i(Y;\mu_j)}{\partial \beta_s}\right]\right\}.$$

Recall that $V_u = \sum_{j=1}^{n} h_u(y_j;\mu_j)/\sqrt{n}$. Under H: $\boldsymbol{\theta} = 0$, since the orthonormality implies $\mathrm{E}_0[h_u] = 0$ and $\mathrm{E}_0[Y_j] = \mu_j$,

$$\frac{\partial \log L}{\partial \theta_u} = V_u\sqrt{n} \text{ for } u = 1, \dots, k, \text{ and}$$

$$\frac{\partial \log L}{\partial \beta_s} = \sum_{j=1}^{n}\{y_j - \mu_{0j}\}x_{js} \text{ for } s = 1, \dots, p.$$

To find the asymptotic covariance matrix we need second order derivatives of log L which in turn require the second order derivatives log C_j. We find, for $u, v = 1, \dots, k$,

$$-\frac{\partial^2 \log C}{\partial \theta_u \partial \theta_v} = \mathrm{cov}[h_u\, h_v],$$

and for $u = 1, \dots, k$, $s = 1, \dots, p$

$$-\frac{\partial^2 \log C}{\partial \theta_u \partial \beta_s} = \frac{\partial \log C}{\partial \beta_s}\mathrm{E}[h_u] + \mathrm{E}\left[\frac{\partial h_u}{\partial \beta_s}\right] +$$

$$\mathrm{E}\left[h_u\left\{\sum_{i=1}^{k}\theta_i\frac{\partial h_i}{\partial \beta_s}\right\}\right] + \frac{1}{\mu}\frac{\partial \mu}{\partial \beta_s}\mathrm{E}[h_u(Y-\mu)].$$

The remaining second order derivatives involve factors that simplify when $\boldsymbol{\theta} = 0$, so they will only be given under this constraint. Again the details are omitted. When $\theta = 0$,

$$-\frac{\partial^2 \log C}{\partial \theta_u \partial \theta_v} = \delta_{uv}, \text{ for } u, v = 1, \dots, k,$$

$$-\frac{\partial^2 \log C}{\partial \theta_u \partial \beta_s} = \mathrm{E}_0\left[\frac{\partial h_u}{\partial \beta_s}\right] + \frac{1}{\mu}\frac{\partial \mu}{\partial \beta_s}\mathrm{E}_0[h_u(Y-\mu)],$$

for $u = 1, \dots, k$, $s = 1, \dots, p$, and

$$-\frac{\partial^2 \log C}{\partial \beta_s \partial \beta_t} = 0, \text{ for } s, t = 1, \dots, p.$$

On differentiating the first order derivatives of log L and using previous results we obtain when $\boldsymbol{\theta} = 0$

$$\frac{\partial^2 \log L}{\partial \theta_u \partial \theta_v} = \sum_{j=1}^{n}\frac{\partial^2 \log C_j}{\partial \theta_u \partial \theta_v} = -n\delta_{uv},\ u, v = 1, \dots, k,$$

$$\frac{\partial^2 \log L}{\partial \theta_u \partial \beta_s} = \sum_{j=1}^{n}\frac{\partial^2 \log C_j}{\partial \theta_u \partial \beta_s} + \sum_{j=1}^{n}\frac{\partial h_u}{\partial \beta_s}$$

$$= \sum_{j=1}^{n}\left\{\frac{\partial h_u}{\partial \beta_s} - \mathrm{E}\left[\frac{\partial h_u}{\partial \beta_s}\right]\right\} - \sum_{j=1}^{n}x_{js}\mathrm{E}[h_u(Y_j-\mu_j)] +$$

terms that are zero when $\boldsymbol{\theta} = 0$,
for $u = 1, \dots, k$, $s = 1, \dots, p$,

and, for $s, t = 1, \dots, p$,

$$\frac{\partial^2 \log L}{\partial \beta_s \partial \beta_t} = \sum_{j=1}^{n}\frac{\partial^2 \log C_j}{\partial \beta_s \partial \beta_t}$$

$$+ \sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}\frac{\partial^2 h_i}{\partial \beta_s \partial \beta_t} - \sum_{j=1}^{n}\mu_j x_{js}x_{jt}$$

$$= -\sum_{j=1}^{n}\mu_j x_{js}x_{jt} + \text{terms that are zero when } \boldsymbol{\theta} = 0.$$

To derive the information matrix we finally take minus the expectation under $\boldsymbol{\theta} = 0$ of these second order derivatives. We find

$$-\mathrm{E}_0[\frac{\partial^2 \log L}{\partial \theta_u \partial \theta_v}] = n\delta_{uv},\ u, v = 1, \dots, k,$$

$$-\mathrm{E}_0[\frac{\partial^2 \log L}{\partial \theta_u \partial \beta_s}] = \sum_{j=1}^{n}x_{js}\mathrm{E}_0[h_u(Y_j-\mu_j)] =$$

$$\sum_{j=1}^{n}x_{js}\mathrm{E}_0[h_u h_1]\sqrt{\mu_{0j}} = \delta_{1u}(\mathbf{X}^T\mathbf{D}^{1/2}\mathbf{1}_n)_s$$

for $u = 1, \dots, k$, $s = 1, \dots, p$, and

$$-\mathrm{E}_0[\frac{\partial^2 \log L}{\partial \beta_s \partial \beta_t}] = \sum_{j=1}^{n}\mu_{0j}x_{js}x_{jt} = (\mathbf{X}^T\mathbf{D}\mathbf{X})_{st},$$

for $s, t = 1, \dots, p$.

This leads to the stated blocks of the partitioned information matrix.