Applied Mathematical Sciences, Vol. 3, 2009, no. 39, 1913 - 1921

A Generalized Quasi-Likelihood Estimation of a Poisson Process

M. S. Youssef

The Higher Technological Institute – 10 th Ramadan City, Egypt
yousefselim@yahoo.com

Abstract

In this paper, generalized quasi-likelihood estimation for a Poisson Process model with a dependent structure is developed incorporating knowledge of skewness and kurtosis. This is a generalization of a similar idea proposed for independent observations by Godambe and Thompson (1989). The generalized quasi-score function is constructed on the basis of non-orthogonal estimating function invoking the method of Durairajan (1992).

Keywords: Quasi-likelihood method; Non-orthogonal estimating functions; Poisson regression

1. Introduction

In this paper, we discuss a class of statistical models called generalized linear models that is a natural generalization of classical linear models**.** Generalized linear models are widely used as a standard tool in modern regression analysis. Successful modeling based on generalized linear models relies on correctly specified model components including the random part and the systematic part. In a classical generalized linear model, the random part requires specification of a distribution from the exponential family. This distribution assumption can be relaxed through the specification of a variance function by Wedderburn's (1974) quasi-likelihood approach. The systematic part of the model includes a linear predictor and a link function. The linear predictor typically is a single index, i.e. a linear combination of the predictors. A single index provides a dimension reduction step. The value of the single index is related to the mean through the link function. Correct specification of link and variance functions are key ingredients for successful statistical modeling of a generalized linear model with quasi-likelihood, in the simple situation where a single linear predictor is indeed sufficient for modeling the relationship between covariates and means. We note that consistency of the regression parameter estimates depends on a correctly specified link function, while efficiency depends on a correctly variance function, the importance of choosing a correct link function. However, the price of misspecifying the variance function is not only loss of efficiency of the regression parameter estimates but also incorrect confidence regions and test results.

Within the framework of stochastic processes, Godambe (1985) established the optimality of certain estimating functions that are linear combinations of orthogonal estimating functions and showed that these optimal estimating equations are extensions of the quasi-likelihood equation for independent observations developed by Wedderburn (1974). Thus, the theory of estimating function, in addition to providing satiating equations has led to its extensions. The concept and technique of quasi-likelihood estimation for independent observations have been extended by Godambe and Thompson (1989) by incorporating possible knowledge of the skewness, kurtosis and higher moments of the underlying distribution and the extended quasi-score function has been developed by these authors. The generality of the extended quasi-likelihood estimation of Godambe and Thompson (1989) is derived from the theory of orthogonal estimating functions due to Godambe (1985). Durairajan (1992) considered non-orthogonal 'basis' of estimating functions and obtained a closed form for the optimal estimating function among those spanned by such a basis. In this paper, the approach of Durairajan is employed to develop generalized quasi-likelihood estimation for a semi-parametric model with dependent observations incorporating knowledge of skewness and kurtosis. William and Durairajan (1999) have developed the quasi-likelihood estimation for a semi-parametric model with a dependent structure by incorporating knowledge of skewness and kurtosis. This is a generalization of a similar idea proposed for independent observations by Godambe and Thompson (1989). The generalized quasi-score function is constructed on the basis of non-orthogonal estimating functions invoking the method of Durairajan (1992). The modeling issue is to identify the way in which the variance increases with the mean. In practice it is sometimes the case, that the relationship between the variance and the mean is (approximately), McCullagh and Nelder, 1989, it is often possible to characterize the first two moments of the response variable with unknown distribution of the form:

$$\mathrm{E}\left(y_i\right)=\mu_i\left(\beta\right)$$

$$\mathrm{var}\left(y_i\right)=\mathrm{var}\left(\varepsilon_i\right)=\phi\,\mathrm{E}\left(y_i\right)^{\theta} \quad , \qquad \theta=0.1.2.3$$

. ϕ is possibly unknown scale parameter or dispersion parameter

. $V(\cdot)$ is the variance has known functional form. The function

- that is, the variance is proportional to a power of the mean. Most common values of θ are the values $0,1,2,3$ which correspond to variance functions associated with normal, Poisson, gamma, and inverse Gaussian distributions respectively. The Box–Cox method can be used for investigating whether the variance is of the form in data – and also for identifying a transformation of data onto a scale where the variance is approximately constant. However, it is not necessary to work with data with constant variance provided that the variance function can be identified. A method for doing this is presented in the following. We calculate the variances and means for each group in data, i.e. for each combination of sample.

However, it is not necessary to work with data with constant variance provided that the variance function can be identified.

However, QL naturally comes into plays an important role in connection with

."Normal-Like" data where variance $\left(y_j\right) = \phi V\left(\mu\right) = \phi\,\mu \quad \left(V\left(\mu\right) = 1\right)$

."Poisson-Like" data where variance $\left(y_j\right) = \phi V\left(\mu\right) = \phi\,\mu \quad \left(V\left(\mu\right) = \mu\right)$

."Gamma-Like" data where variance$\left(y_j\right) = \phi V\left(\mu\right) = \phi\,\mu^2$, $\left(V\left(\mu\right) = \mu^2\right)$

This means that if it can be justified that

.The variance is constant as a function of the mean, then one can work with the Gaussian variance function-even if data are not normally distributed.

.The variance is proportional the squared mean, then one can work with the Gamma variance function-even if data not gamma distributed.

. If $\phi = 1$then variance equals the mean, then one can work with the Poisson distributed. **In Section 2**, used non-orthogonal estimating functions. **In Section 3** the semi-parametric model under consideration is described and the generalized quasi-score function is derived. Application of the new theoretical results will bee discussed **in Section 4**.

2. Non-orthogonal and estimating functions

Let $\mathrm{X} = \{x\}$ be an abstract sample space and $\Im = \{F\}$ be a class of distribution function on$\mathrm{X}$. Let $\theta = \left(\theta_1, \cdots, \theta_r, \cdots, \theta_m\right)'$ be a vector parameter with real components defined on $\Im$ such that$\{\theta\} \equiv \Omega$. Let further h_j, $j = 1, \cdots, k$, with arbitrary$k$, be real function on $\mathrm{X} \times \Omega$ such that

$$\mathrm{E}_F\left\{h_j\left(x, \theta(F)\right) \middle| \mathrm{X}_j\right\} = 0\,, F \in \Im, \tag{2.1}$$

where $\mathrm{E}_F\left\{\cdot \middle| \mathrm{X}_j\right\}$ is the expectation under$F$, conditional on $\mathrm{X}_j, \mathrm{X}_j$being a specified partition (or technically a σ-field generated by a partition) of$\mathrm{X}$, $j = 1, \cdots, k$. For simplicity we use the following notation:

$$\mathrm{E}_F\left\{\cdot \,\middle|\, \mathrm{X}_j\right\} \equiv \mathrm{E}_{(j)}\{\cdot\}.$$

Note. Our theory does not require that the functionh_j, $j = 1, \cdots, k$, satisfying (2.1) be exhaustive. The choice of specific function h_j would be determined by the underlying statistical problem. This will be clear from the subsequent applications. To estimate θ on the basis of an observation x we consider the class of estimating functions

$$\wp = \{g\}, g = \left(g_1, \cdots, g_r, \cdots, g_m\right)$$

where

$$g_r = \sum_{j=1}^{k} h_j \, a_{rj} \tag{2.2}$$

a_{rj} being some real function on $\mathrm{X} \times \Omega$ which is measurable with respect to the partition X_j in (2.1), for $j = 1, \cdots, k$. And $r = 1, \cdots, m$. An estimate of θ based on the estimating function g is obtained by solving the estimating equation
$g\ (x, \theta) = 0$ for the observed x .
Let $S(F) = (\sigma_{ij}(F))$ be non-singular for every $F \in \Im$ with

$$\sigma_{ij}(F) = \mathrm{E}_{ij}\left[h_i(x,\theta).h_j(x,\theta)\right],\ i, j = 1, \cdots, k\ , \tag{2.3}$$

where $\mathrm{E}_{ij}(\cdot)$ denotes $\mathrm{E}_F\left\{\cdot \middle| \sigma(\mathrm{X}_i \cup \mathrm{X}_j)\right\}$. Also, let $H(F) = (h_{ij}(F))$ be a $m \times k$ matrix with

$$h_{ij}(F) = \mathrm{E}_j\left(\partial h_j / \partial \theta_i\right),\ j = 1, \cdots, k\ \ i = 1, \cdots, m . \tag{2.4}$$

Denoting $h = (h_1, \cdots h_k)'$, the estimating function
$g^* = H\,S^{-1}h$ is optimal . (2.5)

3. The generalized quasi-score function

To apply the theory of generalized linear models, we replace the abstract sample space $\mathrm{X} = \{x\}$ by R^n, Let $x = (x_1, \cdots, x_n)$ denote the observations from a process with sample space R^n and $f = \{F\}$ be a family of distributions on R^n and $\theta = (\theta_1, \cdots, \theta_m)$ be a parameter defined on f . Let $X_i = \sigma\ (x_1, \cdots, x_{i-1}),\ i = 1, 2, \cdots, n$ be the σ-fields on R^n with X_1 being the trivial σ-field.

$$\mathrm{E}_F\ (x_i) = \mu_i\ \{\theta(F)\} \text{ and } \mathrm{E}_F\ [x_i - \mu_i\ \{\theta(F)\}]^2 = \phi(F) V_i\ \{\theta(F)\} \tag{3.1}$$

for all $F \in \Im'$ μ_i, V_i, $i = 1, \cdots, n$, and ϕ are specified real function of the indicated variables, $\theta = (\theta_1, \cdots, \theta_r, \cdots, \theta_m)$ being as before the vector parameter with real components defined on $\Im'$. The usual setup of generalized linear models (McCullagh and Nelder, 1983) relates to ours as follows. In (3.1), the 'link' function, that is the specified dependence of μ_i on a linear combination of $\theta_1, \cdots, \theta_m$, is not assumed or emphasized. Instead, we assume μ_i to be any specified function of $\theta = (\theta_1, \cdots, \theta_r, \cdots, \theta_n)$. Similarly, in contrast with the usual setup in (3.1), we do not assume the function V_i to depend on θ only through μ_i . The dispersion parameter ϕ in (3.1) is allowed to depend on F , but is functionally independent of θ. The case 'ϕ known' is the case where $\phi(F) = \phi_0$, a known number, for all $F \in \Im'$. This generality is introduced for mathematical clarity and extended application, and of course our results apply directly to the usual setup.. Now in addition to the relationships between means and variances given by (3.1), further suppose that

$$\gamma_{1i} = \mathrm{E}_i\left\{\frac{x_i - \mu_i}{\mathrm{E}_i\left[(x_i - \mu_i)^2\right]^{\frac{1}{2}}}\right\}^3 \quad \text{and} \quad \gamma_{2i} = \mathrm{E}_i\left\{\frac{x_i - \mu_i}{\mathrm{E}_i\left[(x_i - \mu_i)^2\right]^{\frac{1}{2}}}\right\}^4 - 3 \tag{3.2}$$

where γ_{1i} and $\gamma_{2i}, i = 1, \cdots, n$, are assumed known and do not depend on $x_1, \cdots, x_{i-1}$. Note that equation (3.1) prescribes that the conditional mean of x_i given $(x_1, \cdots, x_{i-1})$ depends on the values of $x_1, \cdots, x_{i-1}$. Apart from θ but the conditional variance of x_i. Does not depend on $x_1, \cdots, x_{i-1}$. Also, equation (3.2) prescribes that the skewness and kurtosis of the conditional distribution of x_i given $(x_1, \cdots, x_{i-1})$ are known constants not depending on the values of $x_1, \cdots, x_{i-1}$. Now, for the above semi-parametric model we develop a generalized quasi-likelihood estimation and the associated notion of generalized quasi-score function by considering (a basis of) non-orthogonal estimating functions using the approach of Durairajan (1992).

If we assume that the series is partly autoregressive and partly moving average, we obtain a quite general quasi-likelihood estimation technique.

Let $x = (x_1, \cdots, x_n)$ denote the observations from a process with sample space R^n and $\tau = \{F\}$ be a family of distribution on R^n and $\theta = (\theta_1, \cdots, \theta_r), \phi = (\phi_{r+1}, \cdots, \phi_m)$ be a parameter defined on τ. Let $\mathrm{X}_i = \sigma\{x_1, \cdots, x_{i-1}\},\ i = 1, \cdots, n$ be the σ-field on R^n with X_1 being the trivial σ-field. Where γ_{2i} and $\gamma_{2i},\ i = 1, \cdots, n$ are assumed known and do not depend on $x_1, \cdots, x_{i-1}$. Note that equation (3.1) prescribes that the conditional mean of x_i given $(x_1, \cdots, x_{i-1})$. Depends on the values of $x_1, \cdots, x_{i-1}$ apart from θ but the conditional variance of x_i does not depend on $x_1, \cdots, x_{i-1}$, equation (3.2) prescribes that the skewness and kurtosis of the conditional distribution of x_i given $(x_1, \cdots, x_{i-1})$ are known constants not depending on the values of $x_1, \cdots, x_{i-1}$. Now, for the above semi-parametric model we develop a generalized quasi-likelihood estimation and the associated notion of generalized quasi-score function by considering (a basis of) non-orthogonal estimating function using the approach of Durairajan (1992). We consider the following estimating function:

$$h_i = h_{1i} = x_i - \mathrm{E}(x_i), \quad h_{n+i} = h_{2i} = h_{1i}^2 - \mathrm{var}(x_i), \quad i = 1, \cdots, n \tag{3.3}$$

and the σ-fields $X_i, i = 1, \cdots, n$, as defined at the beginning of the section and $X_{n+j} = X_j,\ j = 1, \cdots, n$. For $i = 1, \cdots, n$

$$\mathrm{E}_i\left(h_{1i}^2\right) = \mathrm{var} = \phi\,\mathrm{V_i}$$

where $h_j,\ j = 1, \ldots, k$, be real functions

$$\mathrm{E}_{i,\,\mathrm{n}+1}\left(h_{1i}h_{2i}\right) = \gamma_{1i}\left(\phi\,\mathrm{V_i}\right)^{\frac{3}{2}}$$

and

$$\mathrm{E}_{n+i}\left(h_{n+i}^2\right) = \left(\gamma_{2i} + 2\right)\phi^2 V_i^2$$

for $i \neq j$

$$\mathrm{E}_{ij}\left(h_{1i}h_{1j}\right) = \mathrm{E}_{ij}\left(h_{ij}h_{2j}\right) = \mathrm{E}_{ij}\left(h_{2i}h_{2j}\right) = 0 \quad , \quad \mathrm{i} \neq \mathrm{j}$$

hence,

$$S = \mathrm{E}_{ij}\left(h_i h_j\right) = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where

$$S_{11} = Diag(\cdots, \phi V_i, \cdots), \qquad S_{12} = S_{21} = Diag\left(\cdots, \gamma_{1i}(\phi V_i)^{\frac{3}{2}}, \cdots\right)$$

and

$$S_{22} = Diag\left(\cdots, (\gamma_{2i}+2)\phi^2 V_i^2, \cdots\right)$$

now,

$$S^{-1} = \begin{pmatrix} Diag\left(\cdots, \dfrac{\gamma_{2i}+2}{\gamma_{2i}+2-\gamma_{1i}^2}\phi^{-1}V_i^{-1}, \cdots\right) & Diag\left(\cdots, \dfrac{-\gamma_{1i}}{\gamma_{2i}+2-\gamma_{1i}^2}\phi^{-\frac{3}{2}}V_i^{-\frac{3}{2}}, \cdots\right) \\ Diag\left(\cdots, \dfrac{-\gamma_{1i}}{\gamma_{2i}+2-\gamma_{1i}^2}\phi^{-\frac{3}{2}}V^{-\frac{3}{2}}\right) & Diag\left(\cdots, \dfrac{1}{\gamma_{2i}+2-\gamma_{1i}^2}\phi^{-2}V_i^{-2}, \cdots\right) \end{pmatrix} \qquad (3.4)$$

and

$$H = \left(\mathrm{E}_j\left(\partial\, h_j / \partial\, \theta_i\right)\right),$$

$$H = \begin{pmatrix} \dfrac{\partial \mu_1}{\partial \theta_1} \ldots - \dfrac{\partial \mu_n}{\partial \theta_1} - \phi \dfrac{\partial V_1}{\partial \theta_1} \ldots - \phi \dfrac{\partial V_n}{\partial \theta_1} \\ \vdots \qquad\qquad\qquad \vdots \\ -\dfrac{\partial \mu_i}{\partial \theta_m} \cdots - \dfrac{\partial \mu_n}{\partial \theta_m} - \phi \dfrac{\partial V_1}{\partial \theta_m} - \phi \dfrac{\partial V_n}{\partial \theta_m} \end{pmatrix} \qquad (3.5)$$

with some computation, we get the r^{th} element of $g^* = H_{m\times 2n}\, S^{-1}_{2n\times 2n}\, h_{2n\times 1} = m \times 1$ as

$$g^* = \left(g_1^*, g_2^*, \cdots, g_n^*\right)$$

$$g_1^* = \sum_{j=1}^{n} a_{1j}^*\, h_{1j} + \sum_{j=1}^{n} a_{1,j+n}^*\, h_{2j}$$

$$g_2^* = \sum_{j=1}^{n} a_{2j}^*\, h_{1j} + \sum_{j=1}^{n} a_{2,j+n}^*\, h_{2j}$$

$$\vdots$$

$$g_m^* = \sum_{j=1}^{n} a_{mj}^*\, h_{1j} + \sum_{j=1}^{n} a_{m,j+n}^*\, h_{2j}$$

$$= -\left\{ \begin{array}{l} \sum_{j=1}^{n} \phi^{-1} V_j^{-1} \left(x_j - \mu_j\right) \left(\frac{\partial \mu_j}{\partial \theta_r} \right) - \phi^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{\gamma_{1j} V_j^{-\frac{1}{2}}}{\gamma_{2j} + 2 - \gamma_{1j}^2} \frac{\partial \mu_j}{\partial \theta_r} \\ \times \left[\phi^{-1} V_j^{-1} \left(x_j - \mu_j\right)^2 - 1 - \gamma_{1j} \phi^{-\frac{1}{2}} V_j^{-1} \left(x_j - \mu_j\right) \right] \\ + \sum_{j=1}^{n} \frac{V_j^{-1}}{\gamma_{2j} + 2 - \gamma_{1j}^2} \frac{\partial V_j}{\partial \theta_r} \left[\phi^{-1} V_j^{-1} \left(x_j - \mu_j\right)^2 - 1 - \gamma_{1j} \phi^{-\frac{1}{2}} V_j^{-\frac{1}{2}} \left(x_j - \mu_j\right) \right] \end{array} \right\} \tag{3.6}$$

suppose in addition, the dispersion parameter ϕ is to be estimated: Then, there is an additional row in H which is equal to $[0,...,0,-V_1,...,V_n]$. the last element of $g^* = HS^{-1}h$ is then given by

$$g_{m+1}^* = g_\phi^* = \phi^{-1} \sum \frac{1}{\gamma_{2j} + 2 - \gamma_{1j}^2} \left[\frac{\left(x_j - \mu_j\right) \gamma_{1j}}{\left(\phi V_j\right)^{\frac{1}{2}}} + 1 - \frac{\left(x_j - \mu_j\right)^2}{\phi V_j} \right] \tag{3.7}$$

if ϕ is known, the optimal estimating equations for $\theta = (\theta_1, \ldots, \theta_m)'$ are given by equation. (3.6) as $g^* = \left(g_1^*,...,g_m^*\right)' = 0$. If ϕ is unknown, then the estimating equations which a re jointly optimal for (θ, ϕ) are given by equation. (3.6) and (3.7) as $g^* = 0$, $g_\phi^* = 0$. In the literature (Wedderburn, 1974), the first term on the right hand side of (3.6) is called the derivative of the quasi-likelihood function; we would call it the quasi-score function. The quasi-likelihood equation is given by 'the quasi-score function=0'.

4. A discrete skeleton of a Poisson process

Let $\{X_r,\ r \geq 0\}$ be a homogeneous Poisson process with parameter λ where X_r denotes the number of occurrences of a certain event upto time 'r'. Let $\{X_0 = 0,\ X_{r1},\ X_{r2},\ \cdots, X_{rn}\}$ be a discrete skeleton of the process on which observations are available, where $\{r_1 < \cdots < r_n\}$ are any fixed but arbitrary points on the time space. That is, the process $\{X_r\}$ is observed at arbitrary time points on the time space $r_1, r_2, \cdots, r_n$.

Let $r_0 = 0$, $Y_0 = 0$, $Y_i = X_{ri}$, $i = 1, \cdots, n$. Then, $Y_i - Y_{i-1}$ follows Poisson distribution with mean $\lambda\{ r_i - r_{i-1} \}$. Here,

$$\mu_j = Y_{j-1} + \lambda\{ r_j - r_{j-1} \} \quad j = 1, \cdots, n \text{ and}$$
$$\text{Variance}\{Y_j \mid Y_1, \cdots, Y_{j-1}\} = \phi V_j(\lambda), \text{ where}$$
$$\phi = 1, \quad V_j(\lambda) = \lambda(r_j - r_{j-1}) \quad j = 1, \cdots, n \text{ also,}$$
$$\mathrm{E}\left[(Y_j - \mu_j)^3 \mid Y_1, \cdots, Y_{j-1}\right] = \lambda(r_j - r_{j-1}) \text{ and}$$
$$\mathrm{E}\left[(Y_j - \mu_j)^4 \mid Y_1, \cdots, Y_{j-1}\right] = 3\lambda^2 (r_j - r_{j-1})^2 + \lambda(r_j - r_{j-1})$$

Then, again in the notation of section 3,

$$\gamma_{1j} = \frac{1}{\sqrt{\lambda(r_j - r_{j-1})}}, \quad \gamma_{2j} = \frac{1}{\lambda(r_j - r_{j-1})}$$

The generalized quasi-score function = 0' of Equs. (3.6). reduce in this case to

$$\frac{\partial \mu_j}{\partial \lambda} = t_j - t_{j-1} = \frac{\partial V_j}{\partial \lambda}$$

Also**,**

$$\gamma_{1j} V_j^{-1/2} = V_j^{-1}$$

The generalized quasi score function is

$$g^* = \frac{Y_n}{\lambda} - t_n = \frac{X_{tn}}{\lambda - t_n}. \tag{3.8}$$

The equation $g^* = 0$ gives $\hat{\lambda} = \frac{X_{tn}}{t_n}$. It may be noted that the (usual) likelihood function of the observables $(x_{t1}, \cdots, x_{tn})$ from a Poisson process is same as the above g^* given in equs. (4.5) and the solution $\hat{\lambda}$ is same as the usual likelihood estimate.

References

[1] Bhapkar, V.P. (1972). On a measure of efficiency in an estimating equation. Sankhya Ser. A **34**, 467-472

[2] Crowder, M. (1987). On linear and quadratic estimating functions. Biometrika **74**, 591-597.

[3**]** Durairajan, T.M. (1992). Optimal estimating function for non-orthogonal model. J. Statist. Plann. Inference **44**, 381-384.

[4] Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. Biometrika 72, 419-428.

[5] Godambe, V.P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation (with discussion). J. Statist. Plann. Inference **22**, 137-172.

[6] McCullagh, P. and Nelder, J.A. (1983). Generalized linear models. London: Chapman and Hall.

[7] McCullagh P and Nelder JA (1989) Generalized linear models. Chapman and Hall, 2nd edn.

[8] Wedderburn, R.W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika **61**, 439-447.

[9] William, M.L. and Durairajan, T.M. (1999). A generalized quasi-likelihood estimation. J. Statist. Plann. Inference **79**, 237-246.

Received: December, 2008