Monografías del Seminario Matemático García de Galdeano **33**, 229–236 (2006)

AN MM-ALGORITHM FOR A CLASS OF OVERDISPERSED REGRESSION MODELS

S. Dossou-Gbété, C. Demétrio and C. C. Kokonendji

Abstract. The aim of the paper is to provide an algorithm for the computation of the regression parameters estimation in the framework of generalized linear model for count data. Regression parameters are estimated through the minimization of the quasi-likelihood and the main feature of that algorithm, which relies on MM method, is not to resort to matrix inversion as in Newton-Raphson algorithm and Fisher-Scoring method.

Keywords: Count data, exponential dispersion models, Hinde-Demétrio models, generalized linear model, minimization, quasi-likelihood, auxiliary function, MM-algorithm.

§1. Introduction

The Poisson models is a linchpin in the count data statistical modelling toolkit. If the counts are observed along with covariates, the generalized linear models is in the core of the modelization of the expected counts with respect to the covariates. The main property of the Poisson distributions is the equality of their means to their variances. It characterizes the Poisson family of discrete distributions amongst the exponential families of count distributions. It frequently occurs Poisson models are not able to handle the variability demonstrated by the data in hand. There are several ways to overcome the lack of fit of the Poisson models. One of them is to use the negative binomial models when there is evidence of overdispersion. The negative binomial models is generally stated by postulating the distribution variance σ^2 is a quadratic function of its mean m, expressed as $\sigma^2 = m + \frac{1}{\phi}m^2$. An alternative way to handle the negative binomial model is through a linear relationship between the variance and the mean $\sigma^2 = \varphi\mu$ where φ is the Fisher index of dispersion [2, 7].

We consider in this paper an alternative to the negative binomial models, when the quadratic function of the mean is not compatible with the variability shown by the data.

§2. Hinde-Demétrio models for overdispersed count data

2.1. The Hinde-Demétrio models for count data ([5])

Let's consider the function V_p defined as $\mu \mapsto V_p(\mu) = \mu + \mu^p$ and indexed by $p \geq 1$. It is shown in [9] that for each $p \geq 1$, the function V_p is the unit variance function of an additive exponential dispersion model ([8]), named Hinde-Demétrio model in [9], with support $S_p = \mathbb{N} + p\mathbb{N}$. A probability distribution belonging in a Hinde-Demétrio model indexed by some p

is characterized by its mean value $m > 0$ and a dispersion parameter ϕ in such a way that its variance σ^2 is linked to the mean by the relation

$$\sigma^2 = m + \phi^{1-p} m^p = \phi V_p\left(\frac{m}{\phi}\right). \tag{1}$$

2.2. The quasi-likelihood for Hinde-Demétrio models

The quasi-likelihood of the mean m and the dispersion parameter ϕ with respect to a distribution belonging in a Hinde-Demétrio model indexed by p is defined as

$$Q_p(m, \phi \mid y) = -\int_y^m \frac{y-u}{u + \phi^{1-p} u^p}\, du.$$

This results in the equation below

$$Q_p(m, \phi \mid y) = -\phi \int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{\frac{y}{\phi} - v}{\phi V_p(v)}\, dv = -\int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{\frac{y}{\phi} - v}{v + v^p}\, dv.$$

A simple algebra yields to

$$\begin{aligned} Q_p(m, \phi \mid y) &= -\frac{y}{\phi} \int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{1}{v + v^p}\, dv + \phi \int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{1}{1 + v^{p-1}}\, dv \\ &= -\frac{y}{\phi} \left\{ \ln\left[\frac{m + \phi^{1-p} m^p}{y + \phi^{1-p} y^p}\right] - \ln\left[\frac{\phi^{p-1} + m^{p-1}}{\phi^{p-1} + y^{p-1}}\right]^{\frac{p}{p-1}} \right\} \\ &\quad + \phi \int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{1}{1 + v^{p-1}}\, dv. \end{aligned}$$

This result allows to state that:

Lemma 1.

$$Q_p(m, \phi \mid y) = -\frac{y}{\phi} \ln\left[\frac{m\left(1 + \phi^{1-p} m^{p-1}\right)^{-\frac{1}{p-1}}}{y\left(1 + \phi^{1-p} y^{p-1}\right)^{-\frac{1}{p-1}}}\right] + \phi \int_{\frac{y}{\phi}}^{\frac{m}{\phi}} \frac{1}{1 + v^{p-1}}\, dv.$$

Moreover Q_p is a convex positive function on the means domain of the Hinde-Demétrio models and

$$Q_p(m, \phi \mid y) = Q_p(\widehat{m}, \phi \mid y) - \frac{y}{\phi} \ln\left[\frac{m\left(1 + \phi^{1-p} m^{p-1}\right)^{-\frac{1}{p-1}}}{\widehat{m}\left(1 + \phi^{1-p} \widehat{m}^{p-1}\right)^{-\frac{1}{p-1}}}\right] + \phi \int_{\frac{\widehat{m}}{\phi}}^{\frac{m}{\phi}} \frac{1}{1 + v^{p-1}}\, dv.$$

By noticing that $\frac{1}{1+v^{p-1}} \leq \frac{1}{v^{p-1}}$ for $v > 0$ and $x \mapsto \ln(x)$ is a concave function on $[0, +\infty[$, one can readily show the lemma below.

Lemma 2. *Let $p \geq 2$. For any m and $\widehat{m}$ belonging in the means domain of the Hinde-Demétrio models*

$$Q_p(m, \phi \mid y) \leq Q_p(\widehat{m}, \phi \mid y) - \frac{y}{\phi} \ln\left(\frac{m}{\widehat{m}}\right) + \frac{y\phi^{-p}\left[\frac{m^{p-1} - \widehat{m}^{p-1}}{p-1}\right]}{1 + \phi^{1-p} \widehat{m}^{p-1}} + \frac{\phi^{1-p}\left[m^{2-p} - \widehat{m}^{2-p}\right]}{2-p}.$$

§3. Estimation method for the generalized linear models

3.1. Generalized linear regression models

Let y_i, $i = 1 : n$, are count data where each response y_i is observed along with a vector $x_i = (x_{ij})_{j=1:k}$ of k covariates values. We assume that the responses are realisations of independant random variables distributed according to distribution from the same Hinde-Demétrio model with known index parameter p and unknown dispersion parameter ϕ. Conditionaly to the covariates vector x_i, the mean of the distribution underlying to the response y_i is related to x_i as a function $m_i(x_i, \beta)$ where β is a vector of k unkown parameters.

In the generalized linear models framework it is assumed that there is a injective function g, defined on the means domain of the distributions underlying the the response y_i, such that

$$g(m(x_i, \beta)) = \sum_{j=1}^{k} x_{ij}\beta_j = {}^t x_i \beta.$$

Since the responses y_i are modelled by means of distributions belonging in a Hinde-Demetrio family indexed by fixed p with the same dispersion parameter ϕ, the quasi-likelihood of the unknown model parameters ϕ and β_j, $j = 1 : k$, with respect to the data (y_i, x_i), $i = 1 : n$, is

$$Q_p\big(\beta, \phi \mid \{(y_i, x_i),\ i = 1 : n\}\big) = \sum_{i=1}^{n} Q_p(m(x_i, \beta), \phi \mid y_i).$$

For an exponential dispersion family of distributions, the minimization of the quasi-likelihood results in the maximization of the log-likelihood. As a consequence of what come above, the quasi-likelihood can be computed without the explicit knowledge of the cumulant function of the Hinde-Demétrio exponential dispersion models. Then a tentative algorithm for the estimation of the means m_i and the dispersion parameter ϕ is as follows:

Algorithm 1. General procedure for regression parameters estimation.

Repeat until convergence within a numerical tolerance:

1. Hold ϕ fixed.
   Minimize $Q_p\big(\beta, \phi \mid \{(y_i, x_i),\ i = 1 : n\}\big)$.
2. Solve the moment equation $\sum_{i=1}^{n} \frac{(y_i - m(x_i,\beta))^2}{m(x_i,\beta) + \phi^{1-p}[m(x_i,\beta)]^p} =$ degree of freedom with respect to ϕ.

End(Repeat)

Using a surrogate objective function at the place of the actual function to optimize proves to be a convenient gateway on the route of designing algorithms, efficient and easy to implement ([1]). Such surrogate objective functions are sometime called auxiliary functions ([4]). Since auxiliary functions will play a key role in solving the minimization of the quasi-likelihoods that we are dealing with in this paper, let's introduce to basic properties of such functions.

3.2. Introducing to function optimization with auxiliary functions

Definition 1. Let L denote a function on a domain $\mathscr{X} \subset \mathbb{R}^p$. An *auxiliary function* for L is a function A defined on $\mathscr{X} \times \mathscr{X}$ for which the following properties hold:

(i) $\forall x,\, x' \in \mathscr{X},\, L(x) - L(x') \leq A(x,\, x')$,

(ii) $\forall x \in \mathscr{X},\, A(x,\, x) = 0$.

Although the auxiliary functions can be defined in many different ways, what should be beared in mind is that they define pointwise upper-bounds of the gap between two values of the function L. Given a value x', if one can find a value x such that $A(x,\, x') \leq 0$ then $L(x) \leq L(x')$ and a sequence could be carried out, a cluster point of which might be a stationary point for the function L.

Lemma 3. *Let A be an auxiliary function for L and $\hat{x} \in \mathscr{X}$. If, for all $x \in \mathscr{X}$, $A(x,\, \hat{x}) \geq 0$ and A is differentiable in a neighbourhood of $\hat{x}$, then $\frac{\partial A}{\partial x}(x, \hat{x})|_{x=\hat{x}} = 0$.*

Proof. Since $A(\hat{x},\, \hat{x}) = 0$ and, $\forall x \in \mathscr{X}$, $A(x,\, \hat{x}) \geq 0$, then $\hat{x} = \arg\min\{A(x,\, \hat{x}),\, x \in \mathscr{X}\}$ and the lemma holds. □

Proposition 4. *Let A be an auxiliary function for the function L and let $(x_t)_{t\in\mathbb{N}}$ denote a sequence such that $x_{t+1} = \arg\min\{A(x,\, x_t),\, x \in \mathscr{X}\}$. The sequence $\{L(x_t)\}_{t\in\mathbb{N}}$ is a non increasing one. Furthermore, if A is differentiable in a neighbourhood of a cluster point $\hat{x}$ of the sequence $(x_t)_{t\in\mathbb{N}}$, $\frac{\partial A}{\partial x}(x, \hat{x})|_{x=\hat{x}} = 0$.*

Proof. $0 = A(x_t,\, x_t) \geq A(x_{t+1},\, x_t) \geq L(x_{t+1}) - L(x_t)$, then $\{L(x_t)\}_{t\in\mathbb{N}}$ is a non increasing sequence.

Let $\widetilde{x}$ be a cluster point of the sequence $(x_t)_{t\in\mathbb{N}}$. $A(x,\, x_t) \geq A(x_{t+1},\, x_t) \geq L(x_{t+1}) - L(x_t)$. Taking the limits and applying the continuity of the functions L and $A_x(u) = A(x,\, u)$ leads to $A(x,\, \hat{x}) \geq 0$ and $\hat{x} = \arg\min\{A(x,\, \hat{x}),\, x \in \mathscr{X}\}$. □

Corollary 5. *Let $(x_t)_{t\in\mathbb{N}}$ denote a sequence such that $x_{t+1} = \arg\min\{A(x,\, x_t),\, x \in \mathscr{X}\}$. If A is differentiable and*

$$\frac{\partial A}{\partial x}(x, x')\Big|_{x=x'} = 0 \Longrightarrow \frac{\partial L}{\partial x}(x)\Big|_{x=x'} = 0,$$

then any cluster point $\hat{x}$ of $(x_t)_{t\in\mathbb{N}}$ is a stationnary of the function L.

Proof. The proof of the statement comes out readily from the fact that $\frac{\partial A}{\partial x}(x, \hat{x})|_{x=\hat{x}} = 0 \Rightarrow \frac{\partial L}{\partial x}(x)|_{x=\hat{x}} = 0$ □

As a consequence of the above corollary, $x_{t+1} = \arg\min\{A(x,\, x_t),\, x \in \mathscr{X}\}$ might give an appropriate update rule as the core of some minimization algorithm of the function L. Note that the proposition 4. remains true as well as corollary 5. if one considers a sequence $(x_t)_{t\in\mathbb{N}}$ such that $A(x_{t+1}, x_t) \leq 0$ instead of $x_{t+1} = \arg\min\{A(x,\, x_t),\, x \in \mathscr{X}\}$.

3.3. Generalized linear regression with concave link function and Hinde Demétrio models

Let's assume the link function g is concave with values in $\mathbb{R}$. Then its reciprocal $h = g^{-1}$ is a convex function and this yields that the functions $\beta \mapsto \frac{[h({}^t x\beta)]^{p-1}}{p-1}$ and $\beta \mapsto -\ln(h({}^t x\beta))$ are convex functions on the parameters domain $\mathbb{R}^k$.

Lemma 6. *Let $(\alpha_j)_{j=1:k}$ be a vector in $\mathbb{R}^k$ with positive entries such that $\sum_{j=1}^{k} \alpha_j = 1$ and $\widehat{\beta} = \left(\widehat{\beta}_j\right)_{j=1:k} \in \mathbb{R}^k$. Then,*

$$\begin{aligned}
\frac{[h({}^t x\beta)]^{p-1}}{p-1} &= \frac{1}{p-1}\left[h\left\{\sum_{j=1}^{k} \alpha_j \frac{x_j}{\alpha_j}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x\widehat{\beta}\right\}\right]^{p-1} \\
&\leq \sum_{j=1}^{k} \frac{\alpha_j}{p-1}\left[h\left\{\frac{x_j}{\alpha_j}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x\widehat{\beta}\right\}\right]^{p-1}, \\
-\ln(h({}^t x\beta)) &= -\ln\left[h\left\{\sum_{j=1}^{k} \alpha_j \frac{x_j}{\alpha_j}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x\widehat{\beta}\right\}\right] \\
&\leq -\sum_{j=1}^{k} \alpha_j \ln\left[h\left\{\frac{x_j}{\alpha_j}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x\widehat{\beta}\right\}\right].
\end{aligned}$$

By invoking the concavity of the function $v \mapsto \frac{v^{2-p}}{2-p}$, the forthcoming lemma is straightforward.

Lemma 7.

$$Q_p\left(\beta, \phi \mid \{(y_i, x_i)\ i = 1:n\}\right) \leq Q_p\left(\widehat{\beta}, \phi \mid \{(y_i, x_i)\ i = 1:n\}\right) + \sum_{i=1}^{n} A_{p,i}\left(\beta, \widehat{\beta}\right),$$

where

$$\begin{aligned}
A_{p,i}\left(\beta, \widehat{\beta}\right) = &-\frac{y_i}{\phi}\sum_{j=1}^{k} \alpha_{ij} \ln\left[h\left\{\frac{x_{ij}}{\alpha_{ij}}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x_i\widehat{\beta}\right\}\right] + \frac{y_i}{\phi}\ln\left(m\left(x_i, \widehat{\beta}\right)\right) \\
&+ \frac{y_i \phi^{-p}}{1 + \phi^{1-p}\left[m\left(x_i, \widehat{\beta}\right)\right]^{1-p}}\left[\sum_{j=1}^{k} \frac{\alpha_{ij}}{p-1}\left[h\left\{\frac{x_{ij}}{\alpha_j}\left(\beta_j - \widehat{\beta}_j\right) + {}^t x_i\widehat{\beta}\right\}\right]^{p-1}\right] \\
&- \frac{y_i \phi^{-p}}{1 + \phi^{1-p}\left[m\left(x_i, \widehat{\beta}\right)\right]^{1-p}} \frac{\left[m\left(x_i, \widehat{\beta}\right)\right]^{1-p}}{p-1} \\
&+ \phi^{1-p}\left[m\left(x_i, \widehat{\beta}\right)\right]^{1-p} \sum_{j=1}^{k} x_{ij}\left(\beta_j - \widehat{\beta}_j\right).
\end{aligned}$$

Proposition 8. *$A_p(\beta,\widehat{\beta}) = \sum_{i=1}^{n} A_{p,i}(\beta,\widehat{\beta})$ is an auxiliary function for the minimization of $Q_p(\beta,\phi)$.*

One can notice that the parameters β_j, $j = 1:k$, are separated in the function A_p. This makes the minimization of A_p with respect to β easy to carry out, without matrix inversion. What comes before suggests that the regression parameters $\beta \in R^k$ and $\phi > 0$ could be computed by running the following algorithm:

Algorithm 2. Algorithm for generalized linear regression parameters estimation in case of Hinde-Demétrio models.

Repeat until convergence within a numerical tolerance:

1. Hold ϕ fixed.
   Repeat until convergence within a numerical tolerance:
   Solve $\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta}) = 0,\ j = 1:k.$
   End(Repeat)
2. Solve the moment equation $\sum_{i=1}^{n} \frac{\left(y_i h({}^t x_i\widehat{\beta})\right)^2}{h\left({}^t x_i\widehat{\beta}\right)+\phi^{1-p}\left[h\left({}^t x_i\widehat{\beta}\right)\right]^p} =$ degree of freedom with respect to ϕ

End(Repeat)

3.3.1. Solving the system of equations $\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta}) = 0,\ j = 1:k$

One can easily show that

$$\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}} = -\frac{1}{\phi}\sum_{i=1}^{n} y_i \frac{x_{ij} h'\{{}^t x_i\widehat{\beta}\}}{h\{{}^t x_i\widehat{\beta}\}} + \sum_{i=1}^{n} \frac{y_i \phi^{-p} x_{ij} h'\{{}^t x_i\widehat{\beta}\} h\{{}^t x_i\widehat{\beta}\}^{p-2}}{1+\phi^{1-p}\left[m(x_i,\widehat{\beta})\right]^{1-p}} + \phi^{1-p}\sum_{i=1}^{n}\left[m(x_i,\widehat{\beta})\right]^{1-p} x_{ij}$$

and

$$\frac{\partial^2 A_p}{\partial \beta_j^2}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}} = -\frac{1}{\phi}\sum_{i=1}^{n} \frac{y_i x_{ij}^2}{\alpha_{ij}} \frac{h''\{{}^t x_i\widehat{\beta}\} h\{{}^t x_i\widehat{\beta}\}}{\left[h\{{}^t x_i\widehat{\beta}\}\right]^2} + \frac{1}{\phi}\sum_{i=1}^{n} \frac{y_i x_{ij}^2}{\alpha_{ij}} \frac{\left[h'\{{}^t x_i\widehat{\beta}\}\right]^2}{\left[h\{{}^t x_i\widehat{\beta}\}\right]^2} + \sum_{i=1}^{n} \frac{y_i \phi^{-p} x_{ij}^2}{\alpha_{ij}} \frac{h''\{{}^t x_i\widehat{\beta}\} h\{{}^t x_i\widehat{\beta}\}^{p-2}}{1+\phi^{1-p}\left[m(x_i,\widehat{\beta})\right]^{1-p}} + (p-2)\sum_{i=1}^{n} \frac{y_i \phi^{-p} x_{ij}^2}{\alpha_{ij}} \frac{\left[h'\{{}^t x_i\widehat{\beta}\}\right]^2 h\{{}^t x_i\widehat{\beta}\}^{p-3}}{1+\phi^{1-p}\left[m(x_i,\widehat{\beta})\right]^{1-p}}.$$

One can solve the equation $\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta})=0$ by using the the following update rule

$$\widehat{\beta}_j^{new}=\widehat{\beta}_j+\frac{\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}}}{\frac{\partial^2 A_p}{\partial \beta_j^2}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}}}.$$

3.3.2. Log-linear regression with Hinde-Demétrio models

We will focus in this section on the *log* link function. Then, it comes that

$$\frac{\partial A_p}{\partial \beta_j}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}}=-\frac{1}{\phi}\sum_{i=1}^{n}y_i x_{ij}+\phi^{-p}\sum_{i=1}^{n}y_i x_{ij}\frac{\left[m\left(x_i,\widehat{\beta}\right)\right]^{p-1}}{1+\phi^{1-p}\left[m\left(x_i,\widehat{\beta}\right)\right]^{1-p}}+\phi^{1-p}\sum_{i=1}^{n}\left[m_i\left(x_i,\widehat{\beta}\right)\right]^{1-p}x_{ij}$$

and

$$\frac{\partial^2 A_p}{\partial \beta_j^2}(\beta,\widehat{\beta})\Big|_{\beta=\widehat{\beta}}=(p-1)\phi^{-p}\sum_{i=1}^{n}\frac{y_i x_{ij}^2}{\alpha_{ij}}\frac{\left[m\left(x_i,\widehat{\beta}\right)\right]^{p-1}}{1+\phi^{1-p}\left[m\left(x_i,\widehat{\beta}\right)\right]^{1-p}}.$$

The computation of the regression parameters is achieved by means of the algorithm below:

Algorithm 3. Algorithm for log-linear regression with Hinde-Demétrio models

Repeat until convergence within a numerical tolerance:

1. Hold ϕ fixed.

   Repeat until convergence within a numerical tolerance:

$$\widehat{\beta}_j^{new}=\widehat{\beta}_j+\frac{-\frac{1}{\phi}\sum_{i=1}^{n}y_i x_{ij}+\phi^{-p}\sum_{i=1}^{n}y_i x_{ij}\frac{\left[m\left(x_i,\widehat{\beta}\right)\right]^{p-1}}{1+\phi^{1-p}\left[m\left(x_i,\widehat{\beta}\right)\right]^{1-p}}}{(p-1)\,\phi^{-p}\sum_{i=1}^{n}\frac{y_i x_{ij}^2}{\alpha_{ij}}\frac{\left[m\left(x_i,\widehat{\beta}\right)\right]^{p-1}}{1+\phi^{1-p}\left[m\left(x_i,\widehat{\beta}\right)\right]^{1-p}}}$$

   End(Repeat)

2. Solve the moment equation $\sum_{i=1}^{n}\frac{\left(y_i-\exp\left({}^t x_i\widehat{\beta}\right)\right)^2}{\exp\left({}^t x_i\widehat{\beta}\right)+\phi^{1-p}\left[\exp\left({}^t x_i\widehat{\beta}\right)\right]^p}=$ degree of freedom with respect to ϕ.

End(Repeat)

§4. Concluding remarks

The use of surrogate objective functions for the optimization of loss functions is gaining interest in computational statistics, extending the EM algorithms that are very popular for missing data problems. This paper aims to emphasis this trend by proposing some computation schemes in the framework of the regression analysis of count data that can help to avoid burden of computation in the regression parameters estimation. This preliminary work should be completed by the study of the rates of convergence of the algorithms and the statistical performances of the parameters estimation.

References

[1] BECKER, M. P., YANG, I., AND LANGE, K. EM algorithm without missing data. *Statistical Methods in Medical Research 6* (1997), 38–54.

[2] CAMERON, A. C., AND TRIVEDI, P. K. *Regression Analysis of Count Data*. Cambridge University Press, 1998.

[3] DE LEEUW, J., AND MICHAILIDES, G. Block relaxation algorithms in statistics. `http://www.stat.ucla.edu/deleew/block.pdf` (1998)

[4] DELLA PIETRA, V., ET AL. Induce features of random fields. *IEEE Transactions on pattern recognition and machine intelligence 19*, 4 (1997), 1–13.

[5] HINDE, J., AND DEMÉTRIO, C. *Overdispersion: Models and estimation*. Associação Brasileira de Estatistica, Sao Paulo, 1998.

[6] HUNTER, D. R., AND LANGE, K. A tutorial on MM algorithms. *American statistician 58*, 1 (2004), 30–37.

[7] JANSAKUL, N., AND HINDE, J. P. Linear mean-variance negative binomial models for analysis of orange tissue-culture data. *Songklanakorin J. Sci. Technol. 26*, 5 (2004), 683–696.

[8] JORGENSEN, B. *Exponential Dispersion Models*. Chapman & Hall, London, 1997.

[9] KOKONENDJI, C., DEMÉTRIO, G. B. C., AND DOSSOU-GBÉTÉ, S. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *SORT 28*, 2 (2004)