General Quadratic Distance Methods for Discrete Distributions Definable Recursively [1]

Andrew Luong* and Louis G. Doray**

*École d'actuariat, Université Laval,
Cité Universitaire, Ste-Foy, Québec, Canada G1K 7P4
**Département de mathématiques et de statistique, Université de Montréal,
C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Key Words and Phrases: quadratic distance, recursive relationship, mixture distribution, weighted distribution, goodness-of-fit, minimum chi-square, iteratively reweighted least-squares, truncation.

ABSTRACT

Quadratic distance (QD) methods for inference and hypothesis testing are developed for discrete distributions definable recursively. The methods are general and applicable to many families of discrete distributions including those with complicated probability mass functions (pmf's). Even if no explicit expression for the pmf of some distributions exists, QD methods are relatively simple to implement: the QD estimator can be computed numerically using a nonlinear least squares method. The asymptotic properties of the QD estimator are studied. Test statistics for goodness-of-fit are formulated and shown to follow asymtotically a chi-square distribution under the null hypothesis. Estimation and model testing are treated in a unified way. Simulation results presented indicate that the QDE protects against a certain form of misspecification of the distribution, which makes the maximum likelihood estimator (MLE) biased, while keeping the QDE unbiased.

1 INTRODUCTION

Discrete distributions for modelling count data have been used in many fields of research, which include biometry, actuarial science and economics.

In addition to useful existing discrete distributions, many new parametric families have been formed by procedures of mixture and procedures of generalized stopped sum. We refer to Johnson et al. (1992) for the details of these

[1]This article benefited from remarks made by Professor J. Garrido during a presentation at the Actuarial Day, Laval University, 27 April 2001, by Professor Doray.

procedures. They also give an extensive review of discrete parametric distributions and their applications in various fields. Furthermore, the weighted procedure as formulated by Patil and Rao (1978) or Rao (1989), leads to weighted distributions; the truncated procedure with many applications can be seen as a special case of the weighted procedure.

As a result, many useful discrete distributions have a complicated pmf. In many situations, they involve various forms of series expansion, as in the cases of the Hermite, Pólya-Aeppli or Poisson-generalized inverse Gaussian distribution (see Johnson et al. (1992)).

For these pmf's, the recursive relationship between successive probabilities has been found useful in computing probabilities recursively; see Johnson et al. (1992), Willmot (1993), Feller (1968) and section 2 for some examples.

In this paper, we will consider parametric families where the relationship between successive probabilities of the pmf can be written in the form

$$p_i = \phi_1(\theta, i)p_{i-1} + \ldots + \phi_r(\theta, i)p_{i-r}, \quad (1.1)$$

where

a) $p_i = P[X = i], \quad i = a + r, a + r + 1, \ldots, m \leq \infty$

b) $\theta = [\theta_1, \ldots, \theta_p]'$ is the vector of parameters, $\theta \in \Theta \subseteq \Re^p$

c) $\phi_1, \ldots, \phi_r$, are functions determined by the parametric family.

This is a recurrence equation of order r, i.e. the probability p_i can be expressed in terms of the r previous probabilities $p_{i-1},\ p_{i-2}, \ldots, p_{i-r}$. For many discrete distributions, the recurrence relationship has a simple form with a small value of r.

In this paper, inference techniques which can be viewed as quadratic distance (QD) methods are developed for parametric families which allow recursive relationship (1.1). QD methods are fairly versatile, tractable and suitable for complicated pmf's, since they are not based on the pmf directly. Furthermore, estimation and model testing are treated in a unified way.

For estimation, the quadratic distance estimator (QDE) is flexible, offering a trade-off between efficiency and robustness, if desired (see section 3). For model testing, the test statistics based on quadratic distance follow a unique chi-square distribution across the composite null hypothesis, which makes the test statistics easy to use. QD methods are similar to the minimum chi-square methods and have the advantage of being numerically simple for complicated pmf's (see section 3.4). Therefore, QD methods can be viewed as alternative methods to classical ones such as the method of moments, the maximum likelihood and the minimum chi-square methods. See Moore (1977, 1978) for discussions on the minimum chi-square methods.

The paper is organized as follows. Section 2 illustrates by means of examples many parametric families which allow a recursive representation. In section 3, we introduce the quadratic distance and derive the asymptotic properties of the QDE. Section 4 gives test statistics based on quadratic distance for testing the goodness-of-fit of parametric families for simple and composite hypothesis. Test statistics similar to the Pearson's chi-square statistics are shown to follow a chi-square distribution asymptotically. Section 5 discusses some advantages of the QDE over the MLE when the parametric families are misspecified due to bias sampling and the sampling scheme is not taken into account; finally, section 6 discusses the numerical procedures for implementing the method.

2 RECURRENCE RELATIONSHIP

2.1 Introduction

Let $X_1, \ldots, X_n$ be a random sample from a common pmf $P[X = i] = p_i$, which belongs to a discrete parametric family defined on $\{a, a+1, \ldots, m\}$, and let $\theta = [\theta_1, \ldots, \theta_p]'$ be the vector of parameters to estimate.

Assume p_i satisfies the homogeneous recurrence equation of order r

$$p_i = \phi_1(\theta, i)p_{i-1} + \ldots + \phi_r(\theta, i)p_{i-r}, \quad \text{for } i = a+r, a+r+1, \ldots, m \leq \infty \tag{2.1}$$

where
i) $\phi_1, \ldots, \phi_r$, the functions determined by the parametric family, are assumed to be differentiable, and
ii) $\theta \in \Theta \subseteq \Re^p$.

The recursive relationship between p_i and $p_{i-1}, \ldots, p_{i-r}$ as given by (2.1) with the functions $\phi_1, \ldots, \phi_r$, together with the values $p_a, \ldots, p_{a+r-1}$, implicitly define the parametric family. For some distributions, when $a = 0$, we can define p_{-1} to be 0 and then assume that (2.1) holds, to calculate p_{r-1}. This has the advantage of decreasing the number of initial values which need to be calculated. The Delaporte distribution with $r = 2$ is an example where the probability p_{-1} can be defined as 0 and this value used with p_0 to obtain p_1, even though the domain of definition of this distribution is non-negative (see Willmot (1993)). Only p_0 needs to be calculated in this case; the other values are obtained from (2.1).

We call relationship (2.1), a recursive relationship of order r. With a relationship of order 2 or less, we will be able to represent most parametric families found in Johnson et al. (1992).

2.2 Examples of order 1

Example 1: Panjer's family
The pmf of many useful discrete distributions such as the Poisson, negative binomial, binomial, logarithmic and ETNB distributions can be represented as a recurrence relationship of order 1,

$$p_i = \left(a + \frac{b}{i}\right) p_{i-1}, \qquad \text{starting at } i = 1 \text{ or } i = 2.$$

For the possible values of the coefficients a and b of members of Panjer's classes of distributions $[a, b, 0]$ and $[a, b, 1]$, see Klugman et al. (1998).
Example 2: Good family
The pmf of the Good family (see Doray and Luong (1997)) satisfies the recurrence formula

$$p_i = e^{\alpha} \left(\frac{i}{i-1}\right)^{\beta} p_{i-1}, \quad \alpha < 0, \ \ \beta \in \Re, \quad i = 2, 3, \ldots.$$

The special case $\alpha = 0, \beta < -1$ gives the zeta distribution (see Doray and Luong (1995)).

Other examples of order 1 include the generalized Yule family (see Prasad (1957)), a form of discrete Pareto family, with $\phi_1(\theta, i) = (\lambda + i - 2)/(\lambda + i + 1)$, or the exponential family (see Lehmann (1983)). The generalized Poisson family (see Johnson et al. (1992)) with $\phi_1(\theta, i) = (e^{-\alpha}/i)\left(\frac{[\alpha+i\lambda]^{i-1}}{[\alpha+(i-1)\lambda]^{i-2}}\right)$, is also an example of a recursive relationship of order 1; note that the Poisson distribution is a special case of this family with $\lambda = 0$.

2.3 Examples of order 2

Example 3: Hermite family
The pmf of the Hermite family satisfies the recurrence formula

$$p_i = \left(\frac{\alpha\beta}{i}\right) p_{i-1} + \left(\frac{\alpha^2}{i}\right) p_{i-2}, \quad 0 < \alpha < \sqrt{2}, \ \ \beta > 0, \quad i = 2, 3, \ldots.$$

Example 4: Pólya-Aeppli distribution
The pmf of the Pólya-Aeppli distribution satisfies the recurrence relationship

$$p_i = \left(\frac{\theta(1-p) + 2p(i-1)}{i}\right) p_{i-1} - \left(\frac{p^2(i-2)}{i}\right) p_{i-2}, \quad \theta > 0, \ \ 0 \le p \le 1, \quad i = 2, 3, \ldots.$$

Example 5: Sichel distribution
The pmf of the Sichel distribution satisfies the recurrence relationship

$$p_i = \frac{2\beta}{\alpha}\left(\frac{\gamma + i - 1}{i}\right) p_{i-1} + \left(\frac{\beta^2}{i(i-1)}\right) p_{i-2}, \quad \alpha > 0, \ \ \beta \ge 0, \ \ -\infty < \gamma < \infty, \quad i = 2, 3, \ldots.$$

The Poisson-inverse gaussian distribution is a special case of the Sichel distribution with $\gamma = -1/2$.

Willmot (1993) gives other examples of distributions whose pmf satisfies the recurrence relationship (2.1) of order 2, for example, the Poisson-Pareto, Poisson-beta or Poisson-inverse gamma distributions.

2.4 Examples of order greater than 2

Certain mixed Poisson distributions, such as the Poisson-Weibull, the Poisson-transformed gamma and the Poisson-transformed beta will yield a recurrence relationship of order r greater than 2 with certain values of one of the parameters. The Poisson-Burr and the Poisson-loglogistic distibutions are special cases of the the Poisson-transformed beta. See Willmot (1993) for the values of the functions $\phi_1(\theta, i), \ldots, \phi_r(\theta, i)$.

Note that for some of these distributions, the order r of the recurrence relationship could itself become a parameter of the mixing distribution. In that case, the parameter should be selected as an integer. We assume in this paper that r is fixed and known; the quadratic distance method developed in this paper does not apply to distributions where the order r is unknown.

Sundt (1992) has shown that any distribution on the range $\{0, 1, \ldots, k\}$ with positive probability at 0, will satisfy a recurrence relationship of order k; this holds even for $k = \infty$. However, the recurrence relationship having the lowest order possible should be used.

2.5 Other examples

If p_i allows a representation of the form (2.1), the weighted distributions with pmf

$$p_i^w = \frac{w_i p_i}{E\{w_i\}}$$

where $w_i > 0$ is the weight function without parameters will also satisfy (2.1) with a recurrence relationship of the same order as the original pmf. Often, p_i^w is more complicated than p_i due to the denominator term. Those weighted distributions have many applications in biometry (see Patil and Rao (1978), Rao (1989) or Johnson et al. (1992)). Truncating a distribution on the left or the right still preserves the recurrence relationship for the pmf; in fact, truncating corresponds to have $w_i = 0$ for the domain being truncated and $w_i = 1$ elsewhere.

3 ASYMPTOTIC PROPERTIES OF QDE

3.1 Quadratic distance

Let n_i represent the number of observations which take the value i in the sample $X_1, \ldots, X_n$, let $\hat{p}_i = n_i/n$ represent the relative frequency and let us define

$$\Phi_i(\theta) = \phi_1(\theta, i)\hat{p}_{i-1} + \ldots + \phi_r(\theta, i)\hat{p}_{i-r}.$$

Using relation (2.1) and fixing a value for k with $k \leq m$ (in practice, the choice of k is made so that $n_a, \ldots, n_k$ are all positive), we then have the following representation

$$\hat{p}_i = \Phi_i(\theta) + u_i, \qquad i = a+r, a+r+1, \ldots, k, \qquad (3.1)$$

where u_i is the random error associated with $\hat{p}_i$. Let N_i be the random variable representing the number of observations equal to i in a sample of size n. Using the fact that $N_i \sim$ Bin (n, p_i), it is easily seen that $E(u_i) = 0$, for any n. The variance and the covariance of the u_i's are given in Proposition 3.1, the proof of which is in appendix A.

Proposition 3.1:

$$Cov(u_i, u_{i+l}) = \begin{cases} (1/n)\sum_{j=0}^{r-l} \phi_{i,j}\phi_{i,j+l}p_{i-j} & \text{for } l = 0, \ldots, r \\ 0 & \text{for } l = r+1, \ldots, \end{cases}$$

where we define

$$\phi_{i,j} = \begin{cases} -1 & \text{for } j = 0 \\ \phi_j(\theta, i) & \text{for } j = 1, \ldots, r. \end{cases}$$

The reader will notice the similarity between the above expression for $Cov(u_i, u_{i+l})$ and the one for $Cov(X_i, X_{i+l})$ in an $MA(r)$ process in time series.

Using matrix notation, let $\hat{p} = [\hat{p}_{a+r}, \ldots, \hat{p}_k]'$, $\Phi(\theta) = [\Phi_{a+r}(\theta), \ldots, \Phi_k(\theta)]'$ and $u = [u_{a+r}, \ldots, u_k]'$. We then have

$$\hat{p} = \Phi(\theta) + u.$$

The mean and the variance-covariance matrix of u are given by $E(u) = 0$ and $\Sigma(\theta)$, a band matrix. Let $\Sigma = \Sigma(\theta_0)$, where θ_0 is the true vector value of the parameter. Note that $\Sigma(\theta)$ can be obtained using the variance-covariance matrix of a multinomial distribution. Also, let us define $\Sigma^*(\theta) = n\Sigma(\theta)$ and $\Sigma^* = \Sigma^*(\theta_0)$. Σ^* differs from Σ only by a known constant multiple n.

The quadratic distance estimator (QDE) $\hat{\theta}$ is defined to be the vector value which minimizes

$$Q(\theta) = [\hat{p} - \Phi(\theta)]'\Sigma^{*-1}[\hat{p} - \Phi(\theta)]$$

$$= u'(\theta)\Sigma^{*-1}u(\theta), \qquad (3.2)$$

where we define $u_i(\theta) = \hat{p}_i - \Phi_i(\theta)$.

Note that there are some analogies with the nonlinear least squares estimator as given by Amemiya (1985), or Seber and Wild (1989) for example. However, since $\Phi_i(\theta)$ involves the random terms $\hat{p}_{i-1}, \hat{p}_{i-2} \ldots, \hat{p}_{i-r}$, results for standard nonlinear least squares methods are not applicable directly.

We shall see that $Q(\theta)$ defines a discrepancy measure between the empirical cumulative distribution function (cdf) F_n and the parametric family F_θ, $d(F_n, F_\theta)$ which is defined implicity by the recursive relation (2.1). Therefore, $\hat{\theta}$ is a minimum distance type estimator. For other types of minimum distance estimator, such as the minimum Cramer-Von Mises distance estimator, see Boos (1981).

Let $h_i(x)$ be the indicator function of the interval $I_i = (i-1, i]$, define

$$e_i(F, F_\theta) = \int_{\infty}^{\infty} h_i dF - \phi_1(\theta, i) \int_{\infty}^{\infty} h_i dF - \ldots - \phi_r(\theta, i) \int_{\infty}^{\infty} h_i dF$$

and note that with $F = F_n$, $e_i(F_n, F_\theta) = u_i(\theta)$.

Consequently, a discrepancy measure between F_n and F_θ is defined by the quadratic distance $d(F_n, F_\theta) = Q(\theta)$.

If $\hat{\Sigma}^{*-1}$ is a consistent estimate of Σ^{*-1}, the estimator defined by minimizing

$$u'(\theta)\hat{\Sigma}^{*-1}u(\theta) \qquad (3.3)$$

is asymptotically equivalent to $\hat{\theta}$; see subsection 3.3 for more discussions.

3.2 Consistency of the QDE

Note that $d(F_n, F_{\theta_0}) \rightarrow^p 0$, so if $e_i(F_n, F_\theta) \rightarrow^p 0$ only at $\theta = \theta_0$, this will imply $d(F_n, F_\theta) \rightarrow^p 0$ only at $\theta = \theta_0$. We assume that the parameter space has an interior point $\hat{\theta}$, which minimizes $d(F_n, F_\theta)$. Therefore $\hat{\theta}$ is a consistent estimator of θ, i.e. $\hat{\theta} \rightarrow^p \theta_0$. Similar results for quadratic distance estimators in different models are given by Luong and Thompson (1987) and Luong (1991).

3.3 Asymptotic normality

If we assume that the parameter space is compact, the minimum attained is not at a boundary point, and under differentiability assumptions conditions similar to those given in Cox and Hinkley (1974), the $\hat{\theta}$ which minimizes $Q(\theta)$ satisfies the following system of estimating equations

$$\frac{\partial \Phi'(\theta)}{\partial \theta} \Sigma^{*-1}[\hat{p} - \Phi(\theta)] = 0 \qquad (3.4)$$

where

$$\frac{\partial \Phi(\theta)}{\partial \theta} = \left(\frac{\partial \Phi_i(\theta)}{\partial \theta_j} \right), \quad i = a + r, \ldots, k \text{ and } j = 1, \ldots, p,$$

with

$$\frac{\partial \Phi_i(\theta)}{\partial \theta_j} = \sum_{l=1}^{i-r} \frac{\partial \phi_{i-l}(\theta, i)}{\partial \theta_j} \hat{p}_l.$$

Let us define the matrix $S = (s_{ij})$, where

$$s_{ij} = E\left(\frac{\partial \Phi_i(\theta)}{\partial \theta_j} \right) = \sum_{l=1}^{i-r} \frac{\partial \phi_{i-l}(\theta, i)}{\partial \theta_j} p_l, \text{ evaluated at } \theta = \theta_0.$$

We then have, under the hypothesis $\theta = \theta_0$,

$$S = \frac{\partial \Phi(\theta)}{\partial \theta} + o_p(1) \qquad (3.5)$$

with $o_p(1)$ denoting an expression converging to 0 in probability.

A Taylor series' expansion of expression (3.4) around $\theta = \theta_0$ and use of (3.5) yield, after simplifications,

$$\sqrt{n} S' \Sigma^{*-1} [\hat{p} - \Phi(\theta_0)] = S \Sigma^{*-1} S \sqrt{n} (\hat{\theta} - \theta_0) + o_p(1) \quad (3.6)$$

or equivalently,

$$\sqrt{n}(\hat{\theta} - \theta_0) = (S' \Sigma^{*-1} S)^{-1} S' \Sigma^{*-1} \sqrt{n} (\hat{p} - \Phi(\theta_0)) + o_p(1). \quad (3.7)$$

Since $\sqrt{n}(\hat{p} - \Phi(\theta_0)) \to^{\mathcal{L}} N(0, \Sigma^*)$, using a multivariate central limit theorem, this implies

$$\sqrt{n}(\hat{\theta} - \theta_0) \to^{\mathcal{L}} N(0, \Sigma_{\hat{\theta}}), \text{ with } \Sigma_{\hat{\theta}} = (S' \Sigma^{*-1} S)^{-1}. \qquad (3.8)$$

Thus, the asymptotic variance-covariance matrix of $\hat{\theta}$ is $Var(\hat{\theta}) = \frac{1}{n}(S'\Sigma^{*-1}S)^{-1}$. From previous proofs, we see that (3.2) and (3.3) yield asymptotically equivalent estimators. In practice, (3.3) is used to obtain the QDE.

A class of quadratic distance type estimators obtained by minimizing $u'(\theta) U u(\theta)$ where U is a symmetric positive definite matrix, produces consistent estimators with asymptotic variance-covariance matrix given by

$$\frac{1}{n}(S'US)^{-1}(S'U\Sigma^* US)(S'US)^{-1}.$$

The most efficient estimator is $\hat{\theta}$, obtained by choosing $U = \Sigma^{*-1}$. The easiest one to obtain is $\tilde{\theta}$, obtained by choosing $U = I$, the identity matrix. Despite the fact that $\tilde{\theta}$ is less efficient, it can be used to estimate Σ^{*-1}, by letting

$\hat{\Sigma}^{*-1} = \Sigma^{*-1}(\tilde{\theta})$. We then can use $\hat{\Sigma}^{*-1}$ to obtain the first iteration for $\hat{\theta}$ and this procedure can be repeated with Σ^{*-1} reestimated at each step; $\hat{\theta}$ is defined as the convergent vector value of the procedure. For discussion on computational procedures, see section 6.

Admittedly, there is some arbitrariness in fixing a value for k. The QDE remains consistent for all choices of values for k. For efficiency sake, we should fix k at a large value or let $k \to \infty$, as the sample size $n \to \infty$. For robustness sake, we might fix $k = k_0$, discarding possible outlier observations at the tail, or values exceeding k_0.

3.4 Efficiency

The QDE is known to have high efficiency for certain parametric families representable with a recurrence equation of order 1 over an infinite range. For Panjer's (a, b) family, Luong and Garrido (1993) proved that the QDE was highly efficient. They also mention that the proof can be generalized to show that the QDE is highly efficient when the functions $\phi_1(\theta, i)$ and $\phi_2(\theta, i)$ in the second order homogeneous recurrence equation (2.1) are linear in all the parameters,

$$p_i = (\alpha_1 u_{1,i} + \ldots + \alpha_j u_{j,i})p_{i-1} + (\beta_1 u_{1,i} + \ldots + \beta_l u_{l,i-1})p_{i-2},$$

where u is a vector of known constants. Members who satisfy this last equation after a suitable reparametrization, are the generalized Waring and hyper-Poisson distributions for $r = 1$, and the Poisson-generalized inverse Gaussian and Poisson-beta distributions for $r = 2$ (see Willmot and Panjer (1987)).

For the zeta and Good family, Doray and Luong (1995, 1997) calculated the efficiency of the QDE and confirmed it was highly efficient over a wide range of parameter values. At this point, we cannot provide a statement in general concerning efficiency, since it depends on the parametric families considered.

3.5 The special case of $r = 1$

A variation of the above general QD method which can lead to simplifications in computations exists for parametric families which allow a recursive relationship of order 1.

With $r = 1$, the recursive relationship (2.1) is reduced to

$$p_i = \phi_1(\theta, i)p_{i-1}, \qquad i = a + 1, \ldots, m. \quad (3.9)$$

Let us define $\phi_i = \phi_1(\theta, i)$. Relation (3.9) is equivalent to

$$p_i/p_{i-1} = \phi_i, \qquad i = a + 1, \ldots, m. \quad (3.10)$$

Choosing a value for k, let $y_i = \hat{p}_i/\hat{p}_{i-1}, \quad i = a+1, \ldots, k$. Let us define the vectors $Y = [y_{a+1}, \ldots, y_k]'$ and $\phi(\theta) = [\phi_{a+1}, \ldots, \phi_k]'$; we then have $Y = \phi(\theta) + \epsilon$, where $\epsilon_i = \hat{p}_i/\hat{p}_{i-1} - \phi_i$ and $\epsilon = [\epsilon_{a+1}, \ldots, \epsilon_k]'$.

Using an argument based on the variance-covariance matrix of a multinomial distribution, we can show that, asymptotically, $E(\epsilon) = 0$ and the variance-covariance matrix of ϵ, $\Sigma(\theta)$, is a tridiagonal matrix equal, with $a = 0$, to

$$\frac{1}{n}\begin{pmatrix} \phi_1^2(\frac{1}{p_0}+\frac{1}{p_1}) & \frac{-1}{p_0} & 0 & \cdots & 0 \\ \frac{-1}{p_0} & \phi_2^2(\frac{1}{p_1}+\frac{1}{p_2}) & \frac{-1}{p_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \frac{-1}{p_{k-2}} & \phi_{k-1}^2(\frac{1}{p_{k-2}}+\frac{1}{p_{k-1}}) & \frac{-1}{p_{k-1}} \\ 0 & \cdots & 0 & \frac{-1}{p_{k-1}} & \phi_k^2(\frac{1}{p_{k-1}}+\frac{1}{p_k}) \end{pmatrix}.$$

The QDE $\hat{\theta}$ can be obtained by minimizing

$$d(F_n, F_\theta) = [Y - \phi(\theta)]'\Sigma^{*-1}[Y - \phi(\theta)]. \qquad (3.11)$$

Clearly, Σ^{*-1} can be replaced by a consistent estimate which is quite easy to obtain in this case, by estimating p_i by $\hat{p}_i = n_i/n$ and ϕ_i by $\hat{\phi}_i = \hat{p}_i/\hat{p}_{i-1}$. Note that ϕ_i does not contain any random element (it is a function only of θ) and $\hat{\theta}$ is a consistent estimator with asymptotic variance-covariance matrix given by $Var(\hat{\theta}) = \frac{1}{n}(S'\Sigma^{*-1}S)^{-1}$, where

$$S = (s_{ij}), \quad \text{with } s_{ij} = \frac{\partial\phi_1(\theta, i)}{\partial\theta_j}, \text{ evaluated at } \theta = \theta_0.$$

For numerical purposes, it might be simpler to consider a recursive relationship which yields a linear function of the parameters after transformation, if one exists. For example, for the Yule distribution, the model linear in ρ,

$$\frac{\hat{p}_i}{\hat{p}_{i+1}} = 1 + \frac{\rho+1}{i} + \epsilon_i, \qquad \rho > 0, \quad i = 1, 2, \ldots$$

would be preferred over the non-linear model

$$\frac{\hat{p}_i}{\hat{p}_{i-1}} = \frac{i-1}{i+\rho} + \epsilon_i, \qquad \rho > 0, \quad i = 2, 3, \ldots.$$

For the Good distribution, a logarithmic transformation of the ratio p_i/p_{i-1} followed by a reparametrization produced the desired linear function in its two parameters (see Doray and Luong (1997))

$$\ln(p_i/p_{i-1}) = \alpha + \beta\ln(i/(i-1)), \qquad i = 2, 3, \ldots.$$

In this case, $Var(\epsilon) = \Sigma(\theta)$ is asymptotically equal to

$$\frac{1}{n}\begin{pmatrix} (\frac{1}{p_1}+\frac{1}{p_2}) & \frac{-1}{p_2} & 0 & 0 & \dots & 0 \\ \frac{-1}{p_2} & (\frac{1}{p_2}+\frac{1}{p_3}) & \frac{-1}{p_3} & 0 & \dots & 0 \\ 0 & \frac{-1}{p_3} & (\frac{1}{p_3}+\frac{1}{p_4}) & \frac{-1}{p_4} & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \frac{-1}{p_{k-2}} & (\frac{1}{p_{k-2}}+\frac{1}{p_{k-1}}) & \frac{-1}{p_{k-1}} \\ 0 & 0 & \dots & 0 & \frac{-1}{p_{k-1}} & (\frac{1}{p_{k-1}}+\frac{1}{p_k}) \end{pmatrix}.$$

4 GOODNESS-OF-FIT TESTS

In this section, the quadratic distance $d(F_n, F_\theta)$ is used naturally for constructing goodness-of-fit test statistics for the simple hypothesis $H_0 : F = F_{\theta_0}$ and for the composite hypothesis $H_0 : F \in \{F_\theta\}$, where the pmf is defined implicitly by the recursive relation. These are omnibus tests against all alternatives. Test statistics constructed will be shown to follow a chi-square distribution asymptotically. The following theorem for quadratic forms will be used; its proof can be found in Moore (1977, 1978), or Rao (1973).

Theorem. Suppose that the random vector Y of dimension p is $N_p(0, \Sigma)$ and C is any $p\times p$ symmetric positive semi-definite matrix; then the quadratic form $Y'CY$ is chi-square distributed with ν degrees of freedom if ΣC is idempotent and trace $(\Sigma C) = \nu$. (The same result holds asymptotically if C is replaced by a consistent estimate $\hat{C}$ and $Y \to^{\mathcal{L}} N_p(0, \Sigma)$).

4.1 Testing the simple hypothesis H_0

To test $H_0 : F = F_{\theta_0}$, the following test statistic

$$D^2(\theta_0) = nd(F_n, F_{\theta_0}) = nu'(\theta_0)\Sigma^{*-1}(\theta_0)u(\theta_0)$$

can be used. Since $\sqrt{n}u'(\theta_0) \to^{\mathcal{L}} N(0, \Sigma^*)$, we then have $D^2(\theta_0) \to^{\mathcal{L}} \chi^2_{k+1-r-a}$. Doray and Huard (2001) have used this statistic to define an overdispersion test of the Poisson hypothesis vs the negative binomial.

4.2 Testing the composite hypothesis H_0

To test $H_0 : F \in \{F_\theta\}, \theta \in \Theta$ (the full parameter space), the following test statistic

$$D^2(\hat{\theta}) = nd(F_n, F_{\hat{\theta}}) = nu'(\hat{\theta})\Sigma^{*-1}(\hat{\theta})u(\hat{\theta})$$

can be used where $\hat{\theta}$ is the QDE obtained by using the same distance. Since $\sqrt{n}u'(\hat{\theta}) = \sqrt{n}u'(\hat{\theta}_0) - \sqrt{n}S(\hat{\theta} - \theta_0) + o_p(1)$ from a Taylor series' expansion and using (3.7), we then have

$$\sqrt{n}u'(\hat{\theta}) = [I - S(S'\Sigma^{*-1}S)^{-1}S'\Sigma^{*-1}]\sqrt{n}u(\theta_0) + o_p(1). \quad (4.1)$$

Since $\Sigma^{*-1}(\hat{\theta}) \to^p \Sigma^{*-1}$, we then have $\sqrt{n}u'(\hat{\theta}) \to^{\mathcal{L}} N(0, \Sigma_1)$ with

$$\Sigma_1 = [I - S(S'\Sigma^{*-1}S)^{-1}S'\Sigma^{*-1}]\Sigma^*[I - S(S'\Sigma^{*-1}S)^{-1}S'\Sigma^{*-1}]$$

which implies $\Sigma_1\Sigma^* = [I - (S'\Sigma^*S)^{-1}S'\Sigma^{*-1}]$, an indempotent matrix with trace $\Sigma_1\Sigma^* = k + 1 - r - a - p$. Consequently, under H_0 and using Theorem 1, $D^2(\hat{\theta}) \to^{\mathcal{L}} \chi^2_{k+1-r-a-p}$.

For the special case $r = 1$, similar test statistics can be constructed based on quadratic distances of the form $[Y - \phi(\theta)]'\Sigma^{*-1}(\theta)[Y - \phi(\theta)]$. Obviously, these test statistics follow a chi-square distribution under the null hypothesis. Analogous results to those of subsections 4.1, 4.2 continue to hold. See Luong and Doray (1996) for an example with the zeta distribution.

5 QDE AND TRUNCATED FAMILIES

For protection against misspecification of the parametric family as regards to truncation, the QDE has clear advantages over the MLE, which is strictly a parametric estimator.

For example, let us assume that the true family, from which the data were generated, is truncated to the right at q, with pmf expressible as

$$p_i^w = \frac{p_i}{\sum_{i \le q} p_i}. \quad (5.1)$$

Then p_i^w satisfies the recursive relationship

$$p_i^w = \phi_1(\theta, i)p_{i-1}^w + \ldots + \phi_r(\theta, i)p_{i-r}^w,$$

i.e. we obtain the same recursive relationship as the one for the untruncated family p_i, given by (2.1). Even though the family is truncated, the recursive relationship remains valid, and the QDE, which is based on frequencies from p_i^w, remains consistent.

Suppose that the analyst does not know that the data come from the truncated distribution (5.1), and postulates for the model he will use, the untruncated distribution. The MLE, based on the wrong model for p_i, would no longer be a consistent estimator for θ, while the QDE remains consistent, even though the wrong distribution (the untruncated one) is used for estimation.

Therefore, the QDE can be considered as a robust semi-parametric estimator, offering protection against misspecification of the parametric family, while the MLE, strictly a parametric estimator, is less robust. The above analysis is also true for truncation on the left, i.e. when only the data greater than a certain value are observed.

Here is an example that will illustrate the problem of misspecification. Suppose a contract stipulates that a policyholder can not make more than a certain number of claims per year (for example, holders of a Canadian Automobile Association card can not make more than 4 calls a year for boosting their car during the long and cold Canadian winter). If the person who analyzes the data is not aware of this clause of the contract, while the true model should be a distribution truncated to the right, he would wrongly assume that the number of claims comes from an untruncated distribution.

To compare the bias of the MLE vs that of the QDE for the Poisson family truncated on the right at q, for different values of q, we did a simulation study. A sample of 100,000 Poisson random variables with parameter $\lambda=$ 1, 2, 3, 4, and 5 were generated in MATHEMATICA (see Table 1 for the observed frequencies); for each value of λ, the observations were successively truncated at $q = 1, 2, 3, 4, 5$ and the parameter λ estimated with the observations $\leq q$, using formulas for the the MLE and the QDE for the non-truncated distribution, a misspecification.

In Table 2, the estimator $\hat{\lambda}_0$ is the MLE $\sum_{i=0}^{q}(in_i) \Big/ \sum_{i=0}^{q} n_i$. The QDE's $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ minimizes the quadratic form $u'(\theta)Uu(\theta)$, with U respectively equal to Σ^{*-1}, $\hat{\Sigma}^{*-1}$ and I, while $\hat{\lambda}_4$, $\hat{\lambda}_5$ and $\hat{\lambda}_6$ minimize the distance $[Y - \phi(\theta)]'U[Y - \phi(\theta)]$, with U respectively equal to Σ^{*-1}, $\hat{\Sigma}^{*-1}$ (the consistent estimate discussed in section (3.5)), and I. To calculate the QDE's, we minimized directly the quadratic forms in MATHEMATICA, with the FindMinimum command and with k set at the value q.

From Table 2, we observe, for the MLE, that as q increases, the bias tends to 0 since less and less values are truncated; also, truncation at a certain q_0 produces a larger bias when the true value of λ is larger, since a larger λ implies larger observations on average and therefore more truncated ones for a fixed q_0. The QDE produces an asymptotically unbiased estimator of λ, while the MLE has a negative bias, for all values of q and λ. The QDE $\hat{\lambda}_5$ should not be used since, out of 25 trials, the algorithm failed to converge 5 times, no matter what initial value was used. This is due to the fact that the covariance between ϵ_i and ϵ_{i+1} for $\hat{\lambda}_5$, $-1/\hat{p}_{i-1}$, becomes very large as i increases and p_i decreases. This is reflected in Table 2, where no local minimum was found for $\hat{\lambda}_5$ when the truncation point q was equal to 3, 4 or 5. This also occured more frequently when the parameter λ was smaller, implying smaller probabilities.

The estimator $\hat{\lambda}_1$ which minimizes (3.2) is among the best ones, having in

Table 1: Observed frequencies of 100,000 Poisson distributions

i	$\lambda =$ 1	2	3	4	5
0	36552	13731	4953	1837	679
1	36762	26718	14777	7383	3325
2	18638	26798	22255	14619	8624
3	6161	18150	22357	19453	13937
4	1531	9225	17101	19522	17477
5	283	3729	10136	15753	17606
6	66	1207	5122	10442	14688
7	7	332	2061	5944	10418
8	0	93	836	2868	6504
9	0	15	293	1379	3621
10	0	1	77	510	1779
11	0	1	24	194	780
12	0	0	5	68	351
13	0	0	3	22	150
14	0	0	0	5	44
15	0	0	0	1	13
16	0	0	0	0	3
17	0	0	0	0	1
18	0	0	0	0	0

Table 2: Absolute biases of MLE & QDE's under misspecified distribution

q	Estimator	$\lambda =$ 1	2	3	4	5
1	$\hat{\lambda}_0$	0.4986	1.3395	2.2510	3.1992	4.1696
1	$\hat{\lambda}_1$	0.0057	0.0542	0.0166	0.0191	0.1031
1	$\hat{\lambda}_2$	0.0057	0.0542	0.0166	0.0191	0.1031
1	$\hat{\lambda}_3$	0.0057	0.0542	0.0166	0.0191	0.1031
1	$\hat{\lambda}_4$	0.0057	0.0542	0.0166	0.0191	0.1031
1	$\hat{\lambda}_5$	0.0057	0.0542	0.0166	0.0191	0.1031
1	$\hat{\lambda}_6$	0.0057	0.0542	0.0166	0.0191	0.1031
2	$\hat{\lambda}_0$	0.1948	0.8057	1.5879	2.4638	3.3708
2	$\hat{\lambda}_1$	0.0093	0.0201	0.0021	0.0226	0.1123
2	$\hat{\lambda}_2$	0.0093	0.0206	0.0020	0.0228	0.1080
2	$\hat{\lambda}_3$	0.0074	0.0249	0.0032	0.0281	0.1458
2	$\hat{\lambda}_4$	0.0094	0.0199	0.0026	0.0240	0.1265
2	$\hat{\lambda}_5$	0.0094	0.0206	0.0025	0.0244	0.1184
2	$\hat{\lambda}_6$	0.0074	0.0422	0.0108	0.0073	0.0450
3	$\hat{\lambda}_0$	0.0570	0.4219	1.0362	1.8061	2.6517
3	$\hat{\lambda}_1$	0.0068	0.0016	0.0079	0.0172	0.0247
3	$\hat{\lambda}_2$	0.0068	0.0206	0.0078	0.0173	0.0308
3	$\hat{\lambda}_3$	0.0071	0.0249	0.0075	0.0164	0.0682
3	$\hat{\lambda}_4$	0.2638	0.0199	0.0048	0.0220	0.0849
3	$\hat{\lambda}_5$	**	0.0206	0.0086	0.0174	0.0428
3	$\hat{\lambda}_6$	0.0061	0.0421	0.0082	0.0060	0.0537
4	$\hat{\lambda}_0$	0.0100	0.1858	0.6086	1.2448	1.9962
4	$\hat{\lambda}_1$	0.0057	0.0069	0.0213	0.0067	0.0212
4	$\hat{\lambda}_2$	0.0057	0.0062	0.0210	0.0068	0.0243
4	$\hat{\lambda}_3$	0.0070	0.0127	0.0174	0.0052	0.0250
4	$\hat{\lambda}_4$	0.0351	0.0036	0.0241	0.0032	0.0353
4	$\hat{\lambda}_5$	0.0233	**	0.0245	0.0051	0.0279
4	$\hat{\lambda}_6$	0.0056	0.0331	0.0058	0.0064	0.0507
5	$\hat{\lambda}_0$	0.0013	0.0650	0.3199	0.7947	1.4261
5	$\hat{\lambda}_1$	0.0046	0.0100	0.0164	0.0037	0.0054
5	$\hat{\lambda}_2$	0.0044	0.0094	0.0161	0.0035	0.0074
5	$\hat{\lambda}_3$	0.0070	0.0124	0.0138	0.0025	0.0039
5	$\hat{\lambda}_4$	0.3720	0.1371	0.0156	0.0155	0.0312
5	$\hat{\lambda}_5$	**	**	**	0.0080	0.0053
5	$\hat{\lambda}_6$	0.0034	0.0316	0.0067	0.0072	0.0483

**: no local minimum

general, a small bias in absolute value. When n is small, it is preferable to minimize distance (3.2) instead of (3.11), since the expression for the variance-covariance matrix Σ^* is exact for any n in (3.2) but only asymptotic for (3.11); alternatively, (3.3) could be minimized, since it is natural to replace Σ^* with an estimate. If it is preferred to minimize distance $[Y - \phi(\theta)]'U[Y - \phi(\theta)]$ instead of (3.2), the estimator $\hat{\lambda}_6$ with $U = I$ is better than $\hat{\lambda}_4$ or $\hat{\lambda}_5$ in general.

6 DISCUSSION

The quadratic distance methods developed in this paper were applied to families whose pmf could be expressed as a homogeneous recursive relationship of order r. Similar results could be obtained for distributions whose pmf satisfies the non-homogeneous recursive relationship of order r,

$$p_i = \phi_0(\theta, i) + \phi_1(\theta, i)p_{i-1} + \ldots + \phi_r(\theta, i)p_{i-r} \qquad (6.1)$$

with the constant term $\phi_0(\theta, i)$ added. Examples of distributions which satisfy (6.1) are the Poisson mixed with a truncated gamma (for $r = 1$) or with a truncated normal distribution (for $r = 2$) (see Willmot(1993)). The resulting non-homogeneous recursive equations can alternatively be expressed as homogeneous recurrence equations of order $r + 1$. For the Poisson-truncated normal distribution, the value $p_{-1} = 0$ can be used with p_0 to calculate p_1, as we explained in section (2.1) for the Delaporte distribution (see Willmot (1993)).

Numerical computations for QD methods can be based on algorithms developed for nonlinear least squares methods drawing on the analogies of their quadratic forms. We refer to Seber and Wild (1989) or Bates and Watts (1988) for Gauss-Newton algorithms with a series of iterated reweighted least squares computations for nonlinear least squares estimation. Statistical packages such as SAS or S-plus provide functions and routines for nonlinear least squares computations. Minimization programs, such as FindMinimum in MATHEMATICA, could also be used, which proved to be very quick in our case.

The QD methods developed in this paper can easily be extended to situations where covariates are present in the model. See Doray and Arsenault (2001) for an example involving the zeta distribution with one covariate.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada, of the Fonds pour la Contribution à la Recherche du Québec, and of the Insurance Chair of Laval University.

REFERENCES

Amemiya, T. (1985). *Advanced econometrics.* Harvard University Press, Cambridge, MA.

Bates, D.M. and Watts (1988). *Nonlinear regression analysis and its applications.* Wiley, New York.

Boos, D.D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association*, **76**, 663-670.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall, London

Doray, L.G. and Arsenault, M. (2001). Estimators of the regression parameters of the zeta distribution, accepted in *Insurance: Mathematics and Economics*, 17p.

Doray, L.G. and Huard, L. (2001). On Some New Goodness-of-Fit Tests for the Poisson Distribution, in: *New Trends in Statistical Modelling, Proceedings of the 16th International Workshop on Statistical Modelling*, Odense, Denmark, 429-435.

Doray, L.G. and Luong, A. (1995). Quadratic distance estimators for the zeta family. *Insurance: Mathematics and Economics*, **16**, 255-260.

Doray, L.G. and Luong, A. (1997). Efficient estimators for the Good family. *Communication in statistics: Simulation and Computation*, **26**, 1075-1088.

Feller, W. (1968). *An introduction to probability theory and its applications*, Vol. 1. Wiley, New York.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate discrete distributions.* Wiley, New York.

Klugman, S.A., Panjer, H.H., Willmot, G.E. (1998). *Loss Models.* Wiley, New-York.

Lehmann, E.L. (1983). *Theory of point estimation.* Wiley, New-York.

Luong, A. and Doray, L.G. (1996). Goodness-of-fit test statistics for the zeta family. *Insurance: Mathematics and Economics*, **19**, 45-53.

Luong, A. and Garrido, J. (1993). Minimum quadratic distance estimation for a parametric family of discrete distributions defined recursively. *Australian Journal of Statistics*, **35**, 59-67.

Luong, A. and Thompson, M.E. (1987). Minimum distance methods based on quadratic distance for transforms. *Canadian Journal of Statistics*, **15**, 239-251.

Luong, A. (1991). Minimum distance methods based on quadratic distances for transforms in simple linear regression models. *Journal of the Royal Statistical Society, Series B*, **53**, 465-471.

Moore, D.S. (1977). Generalized inverses, Wald's method and the construction of chi-square test of fit. *Journal of the American Statistical Association*, **72**, 131-137.

Moore, D.S. (1978). *Chi-square tests.* Studies in statistics, vol. 19, Mathematical Association of America.

Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size biased sampling with application to wild life populations and human families. *Biometrics*, **34**, 179-189.

Prasad, A. (1957). A new discrete distribution. *Sankhya*, **16**, 353-354.

Rao, C.R. (1973). *Linear statistical inference and its applications.* Wiley, New York.

Rao, C.R. (1989). *Statistics and truth.* SECS. State College, Pennsylvania.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear regression.* Wiley, New York.

Sundt, B. (1992). On some extensions of Panjer's Class of Counting Distributions. *Astin Bulletin*, **22**, 61- 80.

Willmot, G.E. (1993). Recursive evaluation of mixed Poisson probabilities and related quantities. *Scandinavian Actuarial Journal*, 114-133.

Willmot, G.E. and Panjer, H.H. (1987). Difference equation approaches in evaluation of compound distributions. *Insurance: Mathematics and Economics*, **6**, 43-56.

APPENDIX A

Proof of Proposition 3.1 :
Since $N_i \sim$ Bin (n, p_i) and $(N_i, N_j) \sim$ Trinomial (n, p_i, p_j), for $i \neq j$, we know that $\mathrm{Var}(\hat{p}_i) = p_i(1-p_i)/n$ and $\mathrm{Cov}(\hat{p}_i, \hat{p}_j) = -p_i p_j/n, \quad i \neq j$. For $\mathrm{Var}(u_i)$, we find for example,

$$\mathrm{Var}\ (u_i) = \ \mathrm{Var}\ [\hat{p}_i - \phi_{i,1}\hat{p}_{i-1} - \ldots - \phi_{i,r}\hat{p}_{i-r}] =$$

$$(1/n)[p_i(1-p_i) + \phi_{i,1}^2 p_{i-1}(1-p_{i-1}) + \ldots + \phi_{i,r}^2 p_{i-r}(1-p_{i-r}) +$$

$$2\phi_{i,1}p_i p_{i-1} + \ldots + 2\phi_{i,r}p_i p_{i-r}] =$$

$(1/n)\sum_{j=0}^{r} \phi_{i,j}^2 p_{i-j}$, where we have defined $\phi_{i,0} = -1$,

since $[p_i^2 + \phi_{i,1}^2 p_{i-1}^2 + \ldots + \phi_{i,r}^2 p_{i-r}^2 - 2\phi_{i,1}p_i p_{i-1} - \ldots - 2\phi_{i,r}p_i p_{i-r}] =$

$$[p_i - \phi_{i,1}p_{i-1} - \ldots - \phi_{i,r}p_{i-r}]^2 = 0.$$

Similarly, $\mathrm{Cov}\ (u_i, u_{i+r}) =$

$$\mathrm{Cov}\ [\hat{p}_i - \phi_{i,1}\hat{p}_{i-1} - \ldots - \phi_{i,r}\hat{p}_{i-r}, \quad \hat{p}_{i+r} - \phi_{i,1}\hat{p}_{i+r-1} - \ldots - \phi_{i,r}\hat{p}_i] =$$

$$(-1/n)[p_i(p_{i+r} - \phi_{i,1}p_{i+r-1} - \ldots - \phi_{i,r}(1-p_i))] -$$

$$(-1/n)[\phi_{i,1}p_{i-1}(-p_{i+r} + \phi_{i,1}p_{i+r-1} + \ldots + \phi_{i,r}p_i)] - \ldots =$$

$$(-1/n)(\phi_{i,r}p_i).$$

Proceed in the same way for the other covariances.

Dr Louis Doray
Département de mathématiques et statistique
Université de Montréal
B.P. 6128, succursale A
Montréal, Québec
Canada
H3C 3J7

Dr Andrew Luong
École d'actuariat
Université Laval
Cité Universitaire
Ste-Foy, Québec
Canada
G1K 7P4