ACTUARIAL RESEARCH CLEARING HOUSE
1990 VOL. 1

Technical Report No. 6/89, October 1989

MINIMUM QUADRATIC DISTANCE ESTIMATION FOR A A PARAMETRIC FAMILY OF DISCRETE DISTRIBUTIONS DEFINED RECURSIVELY

Andrew Luong and José Garrido

MINIMUM QUADRATIC DISTANCE ESTIMATION FOR A PARAMETRIC FAMILY OF DISCRETE DISTRIBUTIONS DEFINED RECURSIVELY*

By Andrew LUONG

Memorial University of Newfoundland

and José GARRIDO

Concordia University

Key words and phrases: Minimum distance estimation, discrete distributions, robust inference, application to compound distributions.

AMS 1985 subject classifications : 62F10, 62F35, 62P05

ABSTRACT

The c.d.f. of a random sum can easily be computed iteratively when the distribution of the number of i.i.d. elements in the sum is itself defined recursively. Classical estimation procedures for such recursive parametric families require specific distributional assumptions (e.g. Poisson, Negative Binomial). The minimum distance estimation procedure proposed here allows for the selection of the best fitting family member without prior distributional assumptions. It is shown that the estimator thus obtained is consistent and that it can be made either robust or asymptotically efficient. Its asymptotic distribution is also derived and an illustration with Automobile Insurance data is included.

---

*Research supported by operating grants from the National Sciences and Engineering Research Council of Canada.

1. Introduction

Let $x_1, \ldots, x_n$ be i.i.d. observations from a discrete parametric family of p.f.'s defined recursively by:

$$p_{i+1} = (\alpha + \frac{\beta}{i+1})\, p_i \quad \text{for} \quad i = a, a+1, \ldots, k-1 \qquad (1)$$

where p_i denotes $\mathcal{P}(X = i)$, $0 \le a < k-1$ and $\underset{\sim}{\theta} = (\alpha, \beta)'$ is the parameter vector of interest. The problem is to estimate $\underset{\sim}{\theta}$ within Θ (i.e. the set of values for which (1) defines a proper p.f.).

Substituting we find that

$$p_i = (\alpha + \frac{\beta}{i}) \ldots (\alpha + \frac{\beta}{a+1})\, p_a = \prod_{j=a+1}^{i} (\alpha + \frac{\beta}{j})\, p_a$$

and

$$p_a = [\, 1 + (\alpha + \frac{\beta}{a+1}) + (\alpha + \frac{\beta}{a+1})(\alpha + \frac{\beta}{a+2}) + \ldots + (\alpha + \frac{\beta}{a+1}) \ldots (\alpha + \frac{\beta}{i})\,]^{-1}$$

$$= [1 + \sum_{i=a+1}^{k} \prod_{j=a+1}^{i} (\alpha + \frac{\beta}{j})\,]^{-1} .$$

Clearly maximum likelihood (m.l.) estimation is not tractable here; even in the case where k is finite, it requires the roots of high degree polynomials. When k is infinite, p_i does not take a closed form, complicating further the use of m.l. estimation.

The study of discrete parametric families defined recursively as in (1) arises naturally in compound distributions. Often the distribution of the random sum will not have a closed form, but can be computed recursively if the number of i.i.d. random variables in the sum follows a p.f. which belongs to (1); see Panjer(1981) and Panjer and Willmot(1982).

The use of a recursive definition also allows for the parametrization of large families of p.f.'s. For example, for k finite, the two parameter family defined in (1) contains the Truncated Geometric ($\beta = 0$) and the Truncated Poisson ($\alpha = 0$) families as special cases. For k infinite it contains the Poisson and Negative Binomial. The problem is to find which member of the recursive family best fits the observed counts $x_1, \ldots, x_n$ without need to specify further their p.f.. In a sense, this resembles the objective of classical procedures testing for overdispersion (here underdispersion could also be tested, but for the generalized family described below).

Such is the motivation to estimate the parameter $\underset{\sim}{\theta}$ of the recursive family with no further restrictions on Θ by an alternate method based on the minimization of quadratic distances. We shall show that:

a) For the case where k is finite, i.e. a parametric family of truncated p.f.'s, the proposed Minimum Quadratic Distance Estimator (Q.D.E.) is as efficient as the difficult-to-compute m.l. estimator.

b) When the range of the parametric family is finite, it is clear that the m.l. estimator is not robust. The Q.D.E., on the other hand, provides a mean for tailorized estimation; by an appropriate choice of the design matrix the estimator can be made robust, at the cost of full efficiency, or fully efficient but not robust.

It is simple to check that the estimation procedure described in the next section generalizes to the much richer parametric family of p.f.'s defined recursively by:

$$p_i = (\alpha_1 u_{1i} + \alpha_2 u_{2i} + \dots + \alpha_j u_{ji}) p_{i-1} + (\beta_1 u_{1,i-1} + \beta_2 u_{2,i-1} + \dots + \beta_l u_{l,i-1}) p_{i-2}$$

for $i = a, a+1, \dots, k-2$, where $\underset{\sim}{\theta} = (\underset{\sim}{\alpha}, \underset{\sim}{\beta})'$ and $\underset{\sim}{u}$ is a vector of known constants. For notational convenience, however, we will only discuss in detail the estimation problem for (1).

Finally, it should be mentioned that methods other than maximum likelihood are given by Brant(1984), Feuerverger & McDunnough(1981, 1984) and Luong & Thompson(1987) when the parametric distribution does not have a closed form but some transform of it has.

2. *The Minimum Quadratic Distance Estimator (Q.D.E.)*

2.1 *Definition of the Q.D.E.*

We first consider the case where k is finite and, without loss of generality, let $a = 0$. Denote by $\hat{p}_i$ the fraction of sample elements $x_1, \dots, x_n$ taking value $i = 0, 1, \dots, k$. We can then write

$$\hat{\rho}_i = (\alpha + \frac{\beta}{i})\,\hat{\rho}_{i-1} + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, k$$

where $E(\varepsilon_i) = 0$; let W_k be a k-dimensional positive definite symmetric matrix of constants. The similarity with the Weighted Least Squares Estimation problem is clear.

The minimum quadratic distance estimator (Q.D.E.), $\hat{\underset{\sim}{\theta}} = (\hat{\alpha}, \hat{\beta})'$, is thus chosen as to minimize

$$Q_n(\underset{\sim}{\theta}) = Z_k'(\underset{\sim}{\theta})W_k Z_k(\underset{\sim}{\theta}) = Z_k'(\alpha,\beta)W_k Z_k(\alpha,\beta) \qquad (2)$$

where

$$Z_k'(\underset{\sim}{\theta}) = Z_k'(\alpha,\beta) = [\,\hat{\rho}_1 - (\alpha + \beta)\,\hat{\rho}_0 \,,\, \ldots \,,\, \hat{\rho}_k - (\alpha + \frac{\beta}{k})\,\hat{\rho}_{k-1}\,]\,.$$

$Q_n(\underset{\sim}{\theta})$ is a quadratic form in α and β . Its minimization reduces to a weighted least squares computation. This should make the Q.D.E. very attractive from a computational point of view. The case of interest is when k is large in comparison to the number of parameters in $\underset{\sim}{\theta}$. Here no restrictions are applied on Θ , the set of admissible values of $\underset{\sim}{\theta}$. For large samples this should not cause a problem since it is shown that the Q.D.E. is consistent. Substituting $\hat{\underset{\sim}{\theta}}$ for $\underset{\sim}{\theta}$ in (1) should still yield a proper p.f.. In small samples, negative estimated ρ_i values could be obtained in some cases if restrictions on $\underset{\sim}{\theta}$ are not imposed in minimizing (2). Standard minimization techniques under constraints are available and hence will not be discussed here explicitly.

2.2 Properties of the Q.D.E.

2.2.1 Consistency

Let $\underset{\sim}{\theta}_0$ be the true value of the parameter $\underset{\sim}{\theta}$. By definition, as $n \to \infty$, $Z_k(\theta_0) \xrightarrow{\mathcal{P}} \underset{\sim}{0}$ which in turn implies that $Q_n(\theta_0) \xrightarrow{\mathcal{P}} 0$. Consistency is therefore guaranteed if we have $Q_n(\theta) \xrightarrow{\mathcal{P}} 0$ at and only at $\underset{\sim}{\theta} = \underset{\sim}{\theta}_0$.

2.2.2 Robustness

We show here that the Q.D.E. defined in (2) is robust in the sense of bounded influence function (see Hampel,1974, 1986). $\hat{\theta}$ can be considered as a statistical functional $\hat{\underset{\sim}{\theta}}(F_n) = [\hat{\alpha}(F_n), \hat{\beta}(F_n)]'$ defined implicitly as a root of the following system of equations:

$$\frac{\delta}{\delta\alpha} Z_k'(\alpha,\beta)W_k Z_k(\alpha,\beta) = 0$$

$$\frac{\delta}{\delta\beta} Z_k'(\alpha,\beta)W_k Z_k(\alpha,\beta) = 0$$

where the vector Z_k in (2) is re-written as

$$Z_k'(\alpha,\beta) = (\int [h_1 - (\alpha + \beta) h_0]dF_n , \dots ,$$
$$\int [h_k - (\alpha + \frac{\beta}{k}) h_{k-1}]dF_n)' \qquad (3)$$

and $h_i(x) = 1$ if $x \epsilon [i,i+1)$ and 0 otherwise. In order to derive the influence function of $\hat{\underset{\sim}{\theta}}$ we need to introduce some more notation. Let F_θ be the c.d.f. of the x_i values under the parameter value $\underset{\sim}{\theta}$ (and $F_0 = F_{\theta_0}$ be the true c.d.f.). If for a fixed value of $x \geq 0$ and any $\lambda \epsilon [0,1]$ we denote by $F_\lambda = (1-\lambda)F_\theta + \lambda\delta_x$, where δ_x is a degenerate c.d.f. at x, then $Z_k'(\alpha,\beta)$ in (3) is a special case of

$$Z'(\alpha,\beta,F_\lambda) = (\int [h_1 - (\alpha + \beta) h_0]dF_\lambda , \dots ,$$
$$\int [h_k - (\alpha + \frac{\beta}{k}) h_{k-1}]dF_\lambda)'$$

where F_n replaces F_λ. Finally, let

$$H_1(\alpha,\beta,\lambda) = \frac{\delta}{\delta\alpha} Z'(\alpha,\beta,F_\lambda)W_k Z(\alpha,\beta,F_\lambda)$$

and
$$H_2(\alpha,\beta,\lambda) = \frac{\delta}{\delta\beta} Z'(\alpha,\beta,F_\lambda)W_k Z(\alpha,\beta,F_\lambda) .$$

The influence function of $\hat{\underset{\sim}{\theta}} = \hat{\underset{\sim}{\theta}}(F_n)$ at a fixed value $x \geq 0$ is thus given by implicit differentiation to be:

$$IC(x) = \begin{bmatrix} \frac{\delta H_1}{\delta\alpha} & \frac{\delta H_1}{\delta\beta} \\ \frac{\delta H_2}{\delta\alpha} & \frac{\delta H_2}{\delta\beta} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\delta H_1}{\delta\lambda} \\ \frac{\delta H_2}{\delta\lambda} \end{bmatrix} \text{ valued at } \lambda = 0 \text{ and } \underset{\sim}{\theta} = \underset{\sim}{\theta}_0,$$

$$= \left. \frac{\delta\hat{\underset{\sim}{\theta}}(F_\lambda)}{\delta\lambda} \right|_{\lambda=0,\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} = \left. \left[\frac{\delta H}{\delta\underset{\sim}{\theta}} \right]^{-1} \left[\frac{\delta H}{\delta\lambda} \right] \right|_{\lambda=0,\underset{\sim}{\theta}=\underset{\sim}{\theta}_0} .$$

If all derivatives are valued at $\lambda = 0$ and $\underset{\sim}{\theta} = \underset{\sim}{\theta}_0 = (\alpha_0,\beta_0)$, we have:

$$\frac{\delta H_1}{\delta\alpha} = \frac{\delta Z'}{\delta\alpha} W_k \frac{\delta Z}{\delta\alpha} , \qquad \frac{\delta H_1}{\delta\beta} = \frac{\delta Z'}{\delta\alpha} W_k \frac{\delta Z}{\delta\beta} ,$$

$$\frac{\delta H_2}{\delta\alpha} = \frac{\delta Z'}{\delta\beta} W_k \frac{\delta Z}{\delta\alpha} , \qquad \frac{\delta H_2}{\delta\beta} = \frac{\delta Z'}{\delta\beta} W_k \frac{\delta Z}{\delta\beta} ,$$

and
$$\frac{\delta H_1}{\delta\lambda} = \frac{\delta Z'}{\delta\alpha} W_k \frac{\delta Z}{\delta\lambda} , \qquad \frac{\delta H_2}{\delta\lambda} = \frac{\delta Z'}{\delta\beta} W_k \frac{\delta Z}{\delta\lambda} .$$

where

$$\frac{\delta Z'}{\delta \alpha} = - [\int h_0 dF_0 \ , \ \ldots \ , \int h_{k-1} dF_0 \ ] = - [\rho_0, \ \rho_1, \ \ldots, \ \rho_{k-1}],$$

$$\frac{\delta Z'}{\delta \beta} = - [\int h_0 dF_0 \ , \int \frac{h_1}{2} dF_0 \ , \ \ldots \ , \int \frac{h_{k-1}}{k} dF_0 \ ]$$

$$= - [\rho_0, \ \frac{\rho_1}{2}, \ \ldots, \ \frac{\rho_{k-1}}{k} \ ] \ ,$$

and

$$\frac{\delta Z'}{\delta \lambda} = [(h_1(x)-\rho_1) - (\alpha_0+\beta_0)(h_0(x)-\rho_0), \ \ldots \ ,$$
$$(h_k(x)-\rho_k) - (\alpha_0+ \frac{\beta_0}{k})(h_{k-1}(x)-\rho_{k-1}) \ ] \ .$$

It is clear that for k finite, $IC(x)$ is a linear combination of $h_i(x)$ values and that, consequently, it remains bounded. The Q.D.E. is therefore robust in the sense of bounded influence function.

2.2.3 Asymptotic normality

The Q.D.E., $\hat{\underset{\sim}{\theta}}$, minimizes $Q_n(\underset{\sim}{\theta})$. Using a Taylor's series expansion we have:

$$\dot{Q}_n(\hat{\underset{\sim}{\theta}}) = \dot{Q}_n(\underset{\sim}{\theta}_0) + \ddot{Q}_n(\underset{\sim}{\theta}_0)(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0) + o_{\mathcal{P}}(n^{-1/2}) \qquad (4)$$

where $\dot{Q}_n(.)$ and $\ddot{Q}_n(.)$ denote the matrices of first and second derivatives, respectively, and $o_{\mathcal{P}}(.)$ stands for a term converging to 0 in probability at the specified rate. Thus,

$$\ddot{Q}_n(\underset{\sim}{\theta}_0)(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0) = - \dot{Q}_n(\underset{\sim}{\theta}_0) + o_{\mathcal{P}}(n^{-1/2})$$

where

$$\dot{Q}_n(\underset{\sim}{\theta}_0) = 2 \frac{\delta}{\delta \underset{\sim}{\theta}} Z_k'(\underset{\sim}{\theta}_0) W_k Z_k(\underset{\sim}{\theta}_0) = 2 [S'(\underset{\sim}{\theta}_0) + o_{\mathcal{P}}(1)] W_k Z_k(\underset{\sim}{\theta}_0)$$

with $S'(\underset{\sim}{\theta}_0) = E[\frac{\delta}{\delta \underset{\sim}{\theta}} Z_k'(\underset{\sim}{\theta}_0)] = \begin{bmatrix} S_1'(\underset{\sim}{\theta}_0) \\ S_2'(\underset{\sim}{\theta}_0) \end{bmatrix}$. Since

$$\ddot{Q}_n(\underset{\sim}{\theta}_0) = 2 \frac{\delta^2}{\delta \underset{\sim}{\theta}^2} Z_k'(\underset{\sim}{\theta}_0) W_k Z_k(\underset{\sim}{\theta}_0) + 2 \frac{\delta}{\delta \underset{\sim}{\theta}} Z_k'(\underset{\sim}{\theta}_0) W_k \frac{\delta}{\delta \underset{\sim}{\theta}} Z_k(\underset{\sim}{\theta}_0)$$

$$= 2 S'(\underset{\sim}{\theta}_0) W_k S(\underset{\sim}{\theta}_0) + o_{\mathcal{P}}(1) \ .$$

Replacing in (4) gives

$$- S'(\underset{\sim}{\theta}_0) W_k \sqrt{n} \ Z_k(\underset{\sim}{\theta}_0) = S'(\underset{\sim}{\theta}_0) \ W_k \ S(\underset{\sim}{\theta}_0) \sqrt{n} \ (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0) + o_p(1) \ .$$

Consequently,

$$\sqrt{n} \ Z_k(\underset{\sim}{\theta}_0) \xrightarrow{\mathcal{L}} N(\underset{\sim}{0} \ , \ \Sigma_0)$$

where $\Sigma_0 = \Sigma(\underset{\sim}{\theta}_0)$ is the variance-covariance matrix of the vector

$$\underset{\sim}{h}(x;\theta_0) = [\ (h_1(x)-p_1) - (\alpha_0+\beta_0)(h_0(x)-p_0),\ \ldots\ ,$$
$$(h_k(x)-p_k) - (\alpha_0+\frac{\beta_0}{k})(h_{k-1}(x)-p_{k-1})\ ]'. \quad (5)$$

For notational convenience let $S_0 = S(\underset{\sim}{\theta}_0)$. We thus have,

$$\sqrt{n}\ (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_0) \xrightarrow{\mathcal{L}} N(\ \underset{\sim}{0}\ ,\ \Sigma_1\ )$$

where $\Sigma_1 = (S_0'W_kS_0)^{-1}\ S_0'W_k\ \Sigma_0\ W_kS_0\ (S_0'W_kS_0)^{-1}$.

The choice $W_k^* = \Sigma_0^{-1}$ is optimal, it reduces Σ_1 to $\Sigma_1^* = (S_0'\Sigma_0^{-1}S_0)^{-1}$.

Clearly W_k^* is a function of $\underset{\sim}{\theta}_0$. An initial estimate, $\tilde{\underset{\sim}{\theta}}$ say, can be used with a two-stage procedure. A quick first estimate is reached by letting $W_k = I_k$ (the k-dimensional identity matrix). The resulting $\tilde{\underset{\sim}{\theta}}$ is then used to compute $\tilde{W}_k^* = \Sigma^{-1}(\tilde{\underset{\sim}{\theta}})$. Provided that $\tilde{W}_k^* \xrightarrow{\mathcal{P}} \Sigma_0^{-1}$, $\tilde{W}_k^*$ yields an estimator asymptotically equivalent to that obtained with W_k^*. It is derived by minimizing $Z_k'(\underset{\sim}{\theta})\tilde{W}_k^*Z_k(\underset{\sim}{\theta})$.

Note that the above procedure also yields robust and consistent estimators.

2.2.4 Efficiency

Let $\psi(x;\underset{\sim}{\theta}_0) = \begin{bmatrix}\psi_1(x;\underset{\sim}{\theta}_0)\\ \psi_2(x;\underset{\sim}{\theta}_0)\end{bmatrix}$ be the score function, where $\psi_1(x;\underset{\sim}{\theta}_0) = \frac{\delta}{\delta\alpha}\ \ln\ p_x\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}$ and $\psi_2(x;\underset{\sim}{\theta}_0) = \frac{\delta}{\delta\beta}\ \ln\ p_x\Big|_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_0}$, and let $I(\underset{\sim}{\theta}_0)$ be Fisher's information matrix.

From the previous section, $(S_0'\Sigma_0^{-1}S_0)$ is the variance-covariance matrix of $-S_0'\Sigma_0^{-1}\underset{\sim}{h}$, where

$$\underset{\sim}{h}_0 = [(h_1(x)-p_1) - (\alpha_0+\beta_0)(h_0(x)-p_0)]$$
$$\vdots$$
$$\underset{\sim}{h}_{k-1} = [(h_k(x)-p_k) - (\alpha_0+\frac{\beta_0}{k})(h_{k-1}(x)-p_{k-1})]\ ,$$

and $\underset{\sim}{h} = \underset{\sim}{h}(x;\underset{\sim}{\theta}_0) = [\ \underset{\sim}{h}_0,\ \ldots\ ,\ \underset{\sim}{h}_{k-1}\ ]'$ as in (5). Let $\varphi = \varphi(x;\underset{\sim}{\theta}_0) = \sum_{j=0}^{k-1} a_j\underset{\sim}{h}_j$ where the a_j are known constants, and consider the problem of choosing $a_0,\ \ldots\ ,\ a_{k-1}$ to minimize

$$\int (\psi_1 - \sum_{j=0}^{k-1} a_j\underset{\sim}{h}_j\ )^2\ dF_0 = \sum_{i=0}^{k-1} [\psi_1(i;\underset{\sim}{\theta}_0) - \varphi(i;\underset{\sim}{\theta}_0)]^2\ p_i\ ,$$

(all functions valued at the true parameter value $\underset{\sim}{\theta}_0$). The solution, say $\underset{\sim}{a}^* = (a_0^*,\ \ldots\ ,\ a_{k-1}^*)'$ is obtained by solving the set of equations,

$$\sum_{i=0}^{k-1} a_i E(\underset{\sim}{h}_i \underset{\sim}{h}_0) = E(\psi_1 \underset{\sim}{h}_0)$$

$$\vdots$$

$$\sum_{i=0}^{k-1} a_i E(\underset{\sim}{h}_i \underset{\sim}{h}_{k-1}) = E(\psi_1 \underset{\sim}{h}_{k-1}) .$$

In matrix notation this reduces to

$$\Sigma_0 \underset{\sim}{a}^* = - S_1 = \begin{bmatrix} E(\psi_1 \underset{\sim}{h}_0) \\ \vdots \\ E(\psi_1 \underset{\sim}{h}_{k-1}) \end{bmatrix} = - \begin{bmatrix} E(\frac{\delta \underset{\sim}{h}_0}{\delta a}) \\ \vdots \\ E(\frac{\delta \underset{\sim}{h}_{k-1}}{\delta a}) \end{bmatrix}$$

which gives $\underset{\sim}{a}^* = - \Sigma_0^{-1} S_1$.

Defining $\varphi_1^* = \underset{\sim}{a}^{*\prime} \underset{\sim}{h} = \sum_{i=0}^{k-1} a_i^* \underset{\sim}{h}_i$, we can view it as the projection of ψ_1 onto the subspace spanned by $\underset{\sim}{h}_0, \ldots , \underset{\sim}{h}_{k-1}$. Similarly, φ^* can also be viewed as the projection of ψ onto the same subspace. Therefore, if the range of the parametric family is finite, ψ must belong to such subspace and $\psi = \varphi^*$. Consequently the Q.D.E. is asymptotically as efficient as the m.l. estimator.

3. The non-truncated case

If the range of the parametric family is infinite, the above discussion still applies if we let $k \to \infty$ as $n \to \infty$ at an appropriate rate, that is $k = o(\sqrt{n})$. The Q.D.E. is then obtained by minimizing the quadratic form,

$$Z_k'(\underset{\sim}{\theta}) \hat{W}_k^* Z_k(\underset{\sim}{\theta})$$

where $\hat{W}_k^*$ is an estimator of $\Sigma^{-1}(\underset{\sim}{\theta}_0)$, and then taking the limit as k tends to infinity.

The consistency of the resulting estimator is established as above. Note that

$$0 \le Z_k'(\hat{\underset{\sim}{\theta}}) \hat{W}_k^* Z_k(\hat{\underset{\sim}{\theta}}) \le Z_k'(\underset{\sim}{\theta}_0) \hat{W}_k^* Z_k(\underset{\sim}{\theta}_0)$$

and since

$$Z_k'(\underset{\sim}{\theta}_0) \hat{W}_k^* Z_k(\underset{\sim}{\theta}_0) - Z_k'(\underset{\sim}{\theta}_0) \Sigma^{-1}(\underset{\sim}{\theta}_0) Z_k(\underset{\sim}{\theta}_0) \xrightarrow{\mathcal{P}} 0$$

it suffices to have $Z_k'(\underset{\sim}{\theta}) \Sigma^{-1}(\underset{\sim}{\theta}) Z_k(\underset{\sim}{\theta}) \xrightarrow{\mathcal{P}} 0$ at and only at $\underset{\sim}{\theta} = \underset{\sim}{\theta}_0$. This is clearly the case here since,

$$E(Z_k'(\underset{\sim}{\theta}_0) \Sigma^{-1}(\underset{\sim}{\theta}_0) Z_k(\underset{\sim}{\theta}_0)) = k/n$$

tends to 0 as $n \rightarrow \infty$, and

$$Var(Z_k'(\theta_0)\Sigma^{-1}(\theta_0)Z_k(\theta_0)) = 2k/n^2$$

also tends to 0 as $n \rightarrow \infty$. Using Chebyshev's inequality mompletes the proof, implying that $\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0$.

The asymptotic normality is also established as in the finite case. Using a Taylor series expansion like in (4) we have

$$\sqrt{n}\ (\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N[\ 0\ ,\ (S'\Sigma^{-1}S)^{-1}]$$

where $S'\Sigma^{-1}S = \lim_{k\to\infty} S_0'\Sigma^{-1}(\theta_0)S_0$, assuming it exists.

Finally, the Q.D.E. also attains full efficiency in the infinite case, for the score function ψ belongs to the (infinite dimensional) linear space spanned by $\{ \mathbf{h}_0, \mathbf{h}_1, \ldots \}$, and therefore, $\psi = \varphi^*$.

Note that if the design matrix W_k^* is truncated to be finite k-dimensional, the above procedure still yields robust and consistent Q.D.E.'s, although sub-efficient.

4. *A numerical example*

The data in Table 1 pertains to an Automobile Insurance portfolio of a Belgian company, observed for the period between July 1, 1975 to June 30, 1976, and reported in Lemaire(1985). Here k stands for the number of claims reported in the observation period, n_k and $\hat{p}_k$ for the observed frequency of k and its corresponding relative frequency, respectively.

Table 1 : Distribution of Number of claims in Portfolio

Observed			Estimated		
k	n_k	$\hat{p}_k$	I	II	III
0	96,978	0.90656	0.90386	0.90658	0.90657
1	9,240	0.08638	0.09136	0.08629	0.08638
2	704	0.00658	0.00462	0.00662	0.00595
3	43	0.00040	0.00016	0.00047	0.00037
4	9	0.00008	0.00000	0.00003	0.00002
≥5	0	0.00000	0.00000	0.00000	0.00070

Lemaire fitted the Poisson (using the notation of section 1, he gets $\hat{\underset{\sim}{\theta}} = (0 , 0.1011)'$) and Negative Binomial distributions ($\hat{\underset{\sim}{\theta}} = (0.05835 , 0.00941)'$) to this data by maximum likelihood estimation; the resulting fitted relative frequencies are reported, respectively, in columns I and II of Table 1. Both of these distributions belong to the family defined in (1), but according to Kolmogorov-Smirnov's criteria, the fit could be improved. Lemaire proceeds by fitting Generalized Geometric (2 parameters) and Mixed Poisson distributions (3 parameters), none of which belong to this recursive family (nor to its multi-parameter generalization), hence the iterative evaluation of his compound sums is not possible.

With the reported $\hat{p}_i$ values we produced an initial QDE of $\underset{\sim}{\theta} = (\alpha\ \beta)'$ in (1) by minimizing Q_4 in (2) with W_4 equal to the identity matrix. This initial estimate, $\tilde{\underset{\sim}{\theta}} = (0.05698\ 0.03830)'$ was then used to compute the optimal $W_4^* = \Sigma^{-1}(\tilde{\underset{\sim}{\theta}})$. Only two such iterations were required to produce a QDE $\hat{\underset{\sim}{\theta}} = (0.05660 , 0.03868)'$ accurate to 6 decimals (the resulting relative frequencies are reported in column III of Table 1).

The example illustrates how relatively simple it is to obtain a QDE within the family defined in (1). See Garrido & Luong(1989) for a more detailed analysis of QD estimation in the multiparameter case (and its applications to acturial inference problems).

5. *References*

Brant, R.F.(1984) "Approximate likelihood and probability methods based on transforms", *Ann. Statist.*, **12**, pp.989-1005.

Feuerverger, A. and Mc Dunnough, P.(1981) "On the efficiency of empirical characteristic function procedures", *J. Roy. Statist. Soc., Ser. B*, **43**, pp.20-27.

Feuerverger, A. and Mc Dunnough, P.(1984) "Statistical transform methods and their efficiency", *Canad. J. Statist.*, **12**, pp.303-317.

Garrido, J. and Luong, A.(1989) "Minimum distance inference for Panjer & Willmot's recursive family of distributions", *to appear*.

Hampel, F.R.(1974) "The influence curve and its role in robust estimation", *J.Amer.Stat.Assoc.*, 69, 346, pp.383-393.

Hampel, F.R. et al.(1986) *Robust statistics: the approach based on influence functions*, Wiley, New-York.

Lemaire, J.(1985) *Automobile Insurance; Actuarial Models*, Kluwer-Nijhoff, Boston.

Luong, A. and Thompson, M.E.(1987) "Minimum distance methods based on quadratic distances for transforms", *Canad. J. Statist.*, 15, pp.239-251.

Panjer, H.H. and Willmot, G.E.(1982) "Recursions for compound distributions", *ASTIN Bulletin*, 13, pp.1-11.

Panjer, H.H.(1981) "Recursive evaluation of a family of compound distributions", *ASTIN Bulletin*, 12, pp.22-26.

List of Technical Reports

1. Z.Khalil, J-P. Dion and B. Dimitrov
   Branching Processes with Pair-Dependent Families
   (April 1989)
2. Y.H. Wang
   A Compound Poisson Convergence Theorem
   (May 1989)
3. Ronald J. Stern and Henry Wolkowicz
   A Note on Generalized Invariant Cones and the Kronecker Canonical Form
   (May 1989)
4. Ronald J. Stern and Henry Wolkowicz
   Invariant Ellipsoidal Cones
   (June 1989)
5. Ronald J. Stern and Henry Wolkowicz
   Exponential Negativity of the Ice Cream Cone
   (July 1989)
6. Andrew Luong and José Garrido
   Minimum Quadratic Distance Estimation for a Parametric Family of Discrete Distributions Defined Recursively
   (October 1989)

Note: Copies of any technical report may be requested from

Yogendra P. Chaubey
Department of Mathematics and Statistics
Concordia University
7141 Sherbrooke Street West
Montreal, P.Q. H4B 1R6
CANADA