

MINIMUM DISTANCE INFERENCE FOR SUNDT'S DISTRIBUTION

Doray L.G. and Haziza A.

Key words: Statistical computing, numerical algorithm, simulations, iterative reweighted least-squares.

COMPSTAT 2004 section: Algorithms.

Abstract: The probability function of a discrete distribution belonging to Sundt's family satisfies a certain recursive relationship of order k. Maximum likelihood estimation of its parameters is difficult since there is no closed-form expression for the probability function. We propose an alternative method to estimate the parameters, based on the construction of a linear model and the minimization of a quadratic distance. The asymptotic properties of these estimators are investigated: asymptotic normality of their distribution, unbiasedness, efficiency.

The quadratic distance estimator (QDE) of the parameters can be calculated by using an iteratively reweighted least-squares algorithm. With simulated data from Sundt's family, we show how to implement this algorithm.

Another advantage of the minimum quadratic distance is that we can construct a test statistic easily computable with the QDE and derive its asymptotic distribution. This enables us to test a simple hypothesis for the parameter values as well as a composite hypothesis leading to a goodness-of-fit test.

1 Properties of Sundt's Distribution

Sundt (1992) has introduced the following family of discrete distributions. A discrete random variable N, taking non-negative values, belongs to Sundt's family if its probability function satisfies the following recursive relationship of order k

$$p_n = \sum_{i=1}^{k} (a_i + b_i/n) p_{n-i}, \quad k \geq 1, \quad p_{-1}, p_{-2}, \ldots = 0, \qquad (1)$$

where $a = (a_1, \ldots, a_k)$ and $b = (b_1, \ldots, b_k)$ are parameter vectors of the distribution such that (1) defines a probability function, parameters that we want to estimate. We will denote the distribution of N by $\mathcal{R}_k[a, b]$. Let $\Psi(s)$ be the probability generating function (pgf) of N. Sundt (1992) has shown that $N \in \mathcal{R}$ if and only if $\frac{d}{ds} \ln \Psi(s)$ can be written as the ratio of two polynomials, the one in the numerator being of degree $\leq k-1$ and the one in the denominator of degree $\leq k$ and with a constant term equal to 1.

Some well-known distributions such as the binomial, Poisson and negative binomial distributions belong to this family with $k = 1$. See Panjer (1981) for the values of the parameters a and b of these distributions, composing Panjer's family. By using the pgf, Sundt (1992) has shown the following results:

1. the sum of two independent random variables $\mathcal{R}_k[a,b]$ and $\mathcal{R}_l[c,d]$ also follows Sundt's distribution, but of order $k+l$.
2. the sum of two independent random variables $\mathcal{R}_k[a,b]$ and $\mathcal{R}_k[a,d]$ follows a $\mathcal{R}_k[a,e]$ distribution, where $e_i = ia_i + b_i + d_i$, for $i = 1,\ldots,k$.
3. the distribution of the m^{th} convolution of independent and identically distributed random variables $\mathcal{R}_k[a,b]$ follows a $\mathcal{R}_k[a,\beta]$ distribution where $\beta_i = (m-1)ia_i + mb_i$, for $i = 1,\ldots,k$.
4. the convolution of two independent random variables $\mathcal{R}_1[a_1,b_1]$ and $\mathcal{R}_1[a_2,b_2]$ follows a $\mathcal{R}_2[(a_1+a_2, a_1a_2),(b_1+b_2,-(a_1b_2+a_2b_1))]$ distribution.

2 Minimum Quadratic Distance Estimation

The maximum likelihood estimates of the parameters for Sundt's family of order k are difficult to compute, because the roots of a polynomial of high degree need to be found and local maxima must be distinguished from the global maximum. For $k = 1$, which is Panjer's family, Luong and Garrido (1993) have shown how to use the recursive relationship to estimate the parameters of the distribution. We will generalize their method to arbirary k. Let us first define the truncated $\mathcal{R}_k$ family.

Definition: A discrete random variable N with domain $0,1,\ldots,w$ belongs to Sundt's truncated family of finite order k if its probability function satisfies the following recursive equation

$$p_n^* = \sum_{i=1}^{k}(a_i + b_i/n)p_{n-i}^*, \quad p_{-1}^* = \ldots = p_{-(k-1)}^* = 0, \quad n = 1,\ldots,w.$$

The theoretical probabilities p_n^* are estimated with the observed frequencies from the sample,

$$\hat{p}_n^* = \frac{f_n}{m}, \quad n = 0,\ldots,w,$$

where f_i is the number of observations equal to i in the sample of size m. That recursive equation, linear in the parameters $a_1,\ldots,a_k,b_1,\ldots b_k$, suggests the following linear regression model

$$\hat{p}_n^* = \sum_{i=1}^{k}(a_i + b_i/n)\hat{p}_{n-i}^* + \epsilon_n, \quad \hat{p}_{-1}^* = \ldots = \hat{p}_{-(k-1)}^* = 0, \quad n = 1,\ldots,w,$$

where ϵ_n is a random error.

Let us define the two vectors $\hat{Y} = (\hat{p}_1^*, \hat{p}_2^*, \ldots, \hat{p}_w^*)'$, $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_w)'$ and matrix $\hat{X} =$

$$\begin{pmatrix} \hat{p}_0^* & \hat{p}_0^* & 0 & 0 & \ldots & 0 & 0 \\ \hat{p}_1^* & \hat{p}_1^*/2 & \hat{p}_0^* & \hat{p}_0^*/2 & \ldots & 0 & 0 \\ \hat{p}_2^* & \hat{p}_2^*/3 & \hat{p}_1^* & \hat{p}_1^*/3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{p}_{w-1}^* & \hat{p}_{w-1}^*/w & \hat{p}_{w-2}^* & \hat{p}_{w-2}^*/w & \ldots & \hat{p}_{w-k}^* & \hat{p}_{w-k}^*/w \end{pmatrix}.$$

In matrix notation, this model can be rewritten as $\hat{Y} = \hat{X}\theta + \epsilon$, where θ is the parameter vector $(a_1, b_1, \ldots, a_k, b_k)'$. If $w > k$, matrix $\hat{X}$ is of full rank with probability 1, as $m \to \infty$, and it tends in probability to its theoretical part, denoted X. We also have $E(\hat{X}) = X$.

Since f_i follows a binomial (m, p_i^*) distribution, it can easily be shown that $E(\epsilon_i) = 0$ for $i = 1, \ldots, w$. Using the fact that $(f_i, f_j), i \neq j$, has a trinomial (m, p_i^*, p_j^*) distribution, $\mathrm{Var}(\epsilon_i)$ and $\mathrm{Cov}(\epsilon_i, \epsilon_j)$ can be obtained after some tedious calculation. Let us denote by Σ_θ the variance-covariance matrix of vector ϵ, and let us define $\Sigma_\theta^* = m\Sigma_\theta$ (see Luong and Doray (2002) for the terms of this matrix).

The minimum quadratic distance estimator (MQDE) of vector θ is the vector value which minimizes the expression $\epsilon'\Sigma_\theta^{*-1}\epsilon$, the solution of which is given by

$$\hat{\theta} = (\hat{X}'\Sigma_\theta^{*-1}\hat{X})^{-1}\hat{X}'\Sigma_\theta^{*-1}\hat{Y};$$

this is not an estimator in the usual sense, since Σ_θ^* is a function of the unknown vector θ. For that reason, an iteratively reweighted least-squares algorithm must be used:
Step 0: Set $i = 0$ and $\hat{\Sigma}_{\hat{\theta}_0}^* = I_w$, where I_w is the identity matrix of dimension w.
Step 1: Compute $\hat{\theta}_{i+1} = (\hat{X}'\hat{\Sigma}_{\hat{\theta}_i}^{*-1}\hat{X})^{-1}\hat{X}'\hat{\Sigma}_{\hat{\theta}_i}^{*-1}\hat{Y}$.
Step 2: Recalculate $\hat{\Sigma}_{\hat{\theta}_{i+1}}^*$
Step 3: Set $i \leftarrow i + 1$.
Go back to step 1 until convergence is attained.

Luong and Garrido (1993) have shown, for $k = 1$, that $\hat{\theta}_i \xrightarrow{p} \theta$ and $\hat{\Sigma}_{\hat{\theta}_i}^{-1} \xrightarrow{p} \Sigma_\theta^{-1}, i = 1, 2, \ldots$, where $\xrightarrow{p}$ denotes convergence in probability. The proof remains valid for $k > 1$. For any value of k, Haziza (1997) has shown the following results:

1. Vector ϵ has an asymptotic normal distribution $N(0, \Sigma_\theta)$, from which it follows that $\sqrt{m}\hat{\theta} \xrightarrow{\mathcal{L}} N(\theta, (X'\Sigma_\theta^{*-1}X)^{-1})$.
2. If $||(X'\Sigma_\theta^{*-1}X)^{-1})|| < \infty$, $\hat{\theta}$ is a consistent estimator of θ.
3. $\hat{\theta}$ is an aymptotically efficient estimator of θ.

Until now, we have assumed that the observations were coming from a truncated $\mathcal{R}_k$ family. If we assume instead that the sample comes from the $\mathcal{R}_k$ family, i.e. we let w tend to infinity, the consistency, asymptotic normality and asymptotic efficiency of $\hat{\theta}$ will remain valid. In practice, we can set $w = A$, for a large value of $A > 0$, and assume that any observation larger than A is an outlier, which is rejected. In this case, the aymptotic efficiency of $\hat{\theta}$ does not hold, but the two other properties still hold. The value of A should be chosen large enough to have the largest possible asymptotic effiency of $\hat{\theta}$.

In the truncated $\mathcal{R}_2$ family, Haziza (1997) has also shown that the MQDE of θ has the robustness property, meaning that that the influence function is bounded (see Hampel (1974) for an exhaustive presentation on the theory of robustness and Hampel (1986) for an intuitive treatment of influence curves).

3 Tests of Hypothesis

In this section, we will generalize tests considered by Doray and Huard (2002) to distinguish between the Poisson and negative binomial distributions to tests applicable to Sundt's family.

Let us assume that the observed data $n_1, \ldots, n_m$ come from the $\mathcal{R}_k$ family truncated at w. To test the null hypothesis that the sample arose from that distribution with parameter vector $\theta_0 = (a_1^0, b_1^0, \ldots, a_k^0, b_k^0)'$, with all the parameter values specified, we calculate the following distance between the empirical and parametric cdf

$$d(F_m, F_{\theta_0}) = m(\hat{Y} - \hat{X}\theta_0)'\Sigma_{\theta_0}^{*-1}(\hat{Y} - \hat{X}\theta_0).$$

Haziza (1997) has shown that, under H_0, as the sample size $m \to \infty$, the asymptotic distribution of $d(F_m, F_{\theta_0})$ is χ^2_w. We will therefore reject the null hypothesis H_0 at the approximate level α if $d(F_m, F_{\theta_0}) > \chi^2_{w;1-\alpha}$ where $\chi^2_{w;1-\alpha}$ is the 100(1–α) percentile of a χ^2_w distribution.

Suppose, on the other hand, that we want to test the null hypothesis H_0 specifying that the data come from a truncated $\mathcal{R}_l$ family, where $l < k$ and l is a positive integer. Let H_0 be

$$H_0 : \theta = (a_1, b_1, \ldots, a_l, b_l, a_{l+1} = 0, b_{l+1} = 0, \ldots, a_k = 0, b_k = 0)',$$

where the first $2l$ parameters of H_0 are unknown.

Let $\tilde{\theta} = (\tilde{a}_1, \tilde{b}_1, \ldots, \tilde{a}_l, \tilde{b}_l, 0, \ldots, 0)'$ be the MQDE of θ obtained by initially setting $a_{l+1} = b_{l+1} = \ldots = a_k = b_k = 0$ and minimizing the distance

$$m(\hat{Y} - \hat{X}\theta)'\Sigma_{\theta}^{*-1}(\hat{Y} - \hat{X}\theta).$$

Under H_0, the distance

$$d(F_m, F_{\tilde{\theta}}) = m(\hat{Y} - \hat{X}\tilde{\theta})'\Sigma_{\tilde{\theta}}^{*-1}(\hat{Y} - \hat{X}\tilde{\theta})$$

follows an asymptotic χ^2 distribution with $w-l$ degrees of freedom (see Haziza (1997) for the proof, based on the fact that $\Sigma^*_{\theta}\Sigma^{*-1}_{\tilde{\theta}}$ is an idempotent matrix with trace equal to $(w-l)$).

The test will therefore consist in rejecting H_0 at level α if $d(F_m, F_{\tilde{\theta}}) > \chi^2_{w-l;1-\alpha}$ where $\chi^2_{w-l;1-\alpha}$ is the 100(1–α) percentile of a χ^2_{w-l} distribution.

4 Numerical Example

In this section, we will illustrate with simulated data, the method developed to calculate the MQDE of the parameter vector for a special case of Sundt's family.

Let us consider the distribution obtained by truncating the convolution of a Poisson distribution with $\lambda = 2$ and a negative binomial distribution with $s = 2$ and $q = 2$. The domain of this truncated distribution is the set $\{0, 1, \ldots, 8\}$ and its theoretical probabilities are given in Table 1. The sample size was set at $m = 15003$ and the observed frequencies also appear in Table 1.

It is well known (see Schröter (1990)) that the probability function of the above convolution will satisfy the recurrence equation

$$p^*_n = (a + b/n)p^*_{n-1} + (c/n)p^*_{n-2}, \quad p^*_{-1} = 0, \quad n = 1, \ldots, 8.$$

Schröter's family is a special case of Sundt's family with $k = 2$ and parameter b_2 set equal to 0.

We obtain the regression model $\hat{Y} = \hat{X}\theta + \epsilon$, where
$\hat{Y} = (0.066, 0.113, 0.149, 0.161, 0.157, 0.132, 0.115, 0.090)'$,
$\theta = (a, b, c)'$

$$\text{and } \hat{X} = \begin{pmatrix} 0.019 & 0.019 & 0 \\ 0.066 & 0.033 & 0.009 \\ 0.113 & 0.038 & 0.022 \\ 0.149 & 0.037 & 0.028 \\ 0.161 & 0.032 & 0.030 \\ 0.157 & 0.026 & 0.027 \\ 0.132 & 0.019 & 0.022 \\ 0.115 & 0.014 & 0.016 \end{pmatrix}.$$

By setting $\hat{\Sigma}^*_{\hat{\theta}_0} = I_8$, we calculate, from step 1 of the algorithm in Section 2, a first estimate for θ, $\hat{\theta}_1 = (0.556, 2.647, -0.677)'$. Applying the iterative algorithm, convergence was attained after only 5 iterations; the MQDE is equal to $\hat{\hat{\theta}} = (0.641, 2.624, -1.127)'$. The estimated variance-covariance matrix of the parameters is equal to

$$\text{Var}(\hat{\theta}) = \begin{pmatrix} 0.0041 & 0.0021 & -0.0258 \\ 0.0021 & 0.0012 & -0.0135 \\ -0.0258 & -0.0135 & 0.1607 \end{pmatrix}.$$

Table 1: Results of the simulation

n	p_n^*	f_n
0	0.0191	279
1	00636.	986
2	0.1145	1691
3	0.1499	2229
4	0.1616	2408
5	0.1540	2357
6	0.1352	1973
7	0.1123	1730
8	0.0898	1350

We can test that the parameter c is significantly different from 0 in the model; the approximate 95% confidence interval for c is $[-1.786, -0.468]$.

References

[1] 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 16^{th} International Workshop on Statistical Modelling, edited by B. Klein and L. Korsholm, Odense, Denmark, 429 – 435.

[2] Hampel, F.R. (1974). *The influence curve and its role in robust estimation*. Journal of the American Statistical Association **69**, 383 – 393.

[3] Hampel, F.R., Rochetti, E.M., Rousseuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley: New York.

[4] Haziza, A. (1997). *Inférence pour la famille de Sundt*, Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal, 77p.

[5] Luong, A. and Doray, L.G. (2002). *General Quadratic Distance Methods for Discrete Distributions Definable Recursively*. Insurance: Mathematics and Economics **30**, 255 – 267.

[6] 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.

[7] Panjer, H.H. (1981). *Recursive Evaluation of a Family of Compound Distributions*. Astin Bulletin **12**, 22 – 26.

[8] Schröter, K.J. (1990). *On a family of counting distributions and recursions for related compound distributions*. Scandinavian Actuarial Journal, 161 – 175.

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

Acknowledgement: The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and of the Insurance Chair of Laval University.
Address: Département de mathématiques et de statistique
Université de Montréal
C.P. 6128, Succursale Centre-ville
Montréal, Qc
Canada
H3C 3J7
E-mail: doray@dms.umontreal.ca