Efficiency of the minimum quadratic distance estimator for the bivariate Poisson distribution

Louis G. Doray, Baba Madji Alhadji Kolo

Efficiency of the minimum quadratic distance estimator for the bivariate Poisson distribution

Louis G. Doray and Baba Madji Alhadji Kolo [1]

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Key Words and Phrases: Bivariate Poisson distribution; quadratic distance estimator; iteratively reweighted least-squares; asymptotic efficiency.

Abstract: We consider the problem of estimating the three parameters of the bivariate Poisson distribution. From the recursive expression for the probability mass function, which is linear in its parameters, we develop the quadratic distance estimator (QDE), which can be computed with an iteratively reweighted least-squares algorithm. The QDE is unbiased, easy to calculate and admits a simple expression for its variance-covariance matrix. We calculate its asymptotic efficiency for various values of the parameters and compare it with that of other estimators obtained from traditional methods (moments, proportion of double zero and even points methods). The QDE has very high asymptotic efficiency for a certain range of parameter values of interest.

Résumé: Nous considérons le problème de l'estimation des trois paramètres de la loi de Poisson bivariée. À partir de la formule de récurrence pour la fonction de probabilité, linéaire en ses paramètres, nous développons l'estimateur de distance quadratique (EDQ) qui peut être calculé avec un algorithme des moindres carrés pondérés itérés. L'EDQ est non-biaisé, facile à calculer et admet une expression simple pour sa matrice de variance-covariance. Nous calculons son efficacité asymptotique pour différentes valeurs des paramètres et la comparons avec celle d'autres estimateurs obtenus avec les méthodes traditionnelles (moments, proportion de doubles zéros et méthode des points pairs). L'EDQ possède une efficacité asymptotique très élevée pour un certain domaine d'intérêt des valeurs des paramètres.

1 INTRODUCTION

Bivariate discrete probability distributions have found applications in actuarial science for example, to model the number of accidents by bus drivers in two non-overlapping periods of time (see Kocherlakota and Kocherlakota (1992)). In this paper, we develop a new method to estimate the parameters of the bivariate Poisson distribution.

[1] The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

2 QUADRATIC DISTANCE ESTIMATOR

Teicher (1954) showed that the probability function of the discrete bivariate Poisson distribution satisfies the two recurrence relationships

$$\begin{aligned} rp_{r,s} &= \lambda_1 p_{r-1,s} + \lambda_3 p_{r-1,s-1} \quad \text{for} \quad r \geq 1, s \geq 1, \\ \text{and } sp_{r,s} &= \lambda_2 p_{r,s-1} + \lambda_3 p_{r-1,s-1} \quad \text{for} \quad r \geq 1, s \geq 1. \end{aligned} \tag{1}$$

For r or s equal to 0, we can use the following equations

$$\begin{aligned} rp_{r,0} &= \lambda_1 p_{r-1,0} \quad \text{for} \quad r > 0, s = 0, \\ sp_{0,s} &= \lambda_2 p_{0,s-1} \quad \text{for} \quad r = 0, s > 0, \\ \text{where } p_{0,0} &= \exp -(\lambda_1 + \lambda_2 + \lambda_3). \end{aligned} \tag{2}$$

We note that the recurrence equations in (1) are linear in all three parameters λ_1, λ_2 and λ_3 and also that $p_{r,s}$ is a linear function of the previous probabilities. Note also that by defining $p_{r,-1}$ and $p_{-1,s}$ to be equal to 0 for any $r, s \geq 0$, the equations (2) become a special case of the equations (1).

Let (x_i, y_i), $i = 1, 2, \dots n$, denote a sample of size n of independent and identically distributed random vectors from the bivariate Poisson distribution. We denote by $f_{r,s}$ the observed frequency of the couple (r, s) for $r = 0, 1, 2, \dots, u$ and $s = 0, 1, 2, \dots, v$, with $\sum_r \sum_s f_{r,s} = n$.

We divide the sample in two groups, the first one containing the (x_i, y_j) for which $i > j$ and the second one with the (x_k, y_l) for which $k \leq l$. We will use the first recurrence relationship of (1) for the first group and the second one for the other group:

$$\begin{aligned} p_{i,j} &= \frac{\lambda_1}{i} p_{i-1,j} + \frac{\lambda_3}{i} p_{i-1,j-1} \quad \text{for} \quad i > j \\ p_{k,l} &= \frac{\lambda_2}{l} p_{k,l-1} + \frac{\lambda_3}{l} p_{k-1,l-1} \quad \text{for} \quad k \leq l. \end{aligned}$$

These linear recurrence relationships suggest the two linear regression models

$$\begin{aligned} \hat{p}_{i,j} &= \frac{\lambda_1}{i} \hat{p}_{i-1,j} + \frac{\lambda_3}{i} \hat{p}_{i-1,j-1} + \epsilon_{i,j} \quad \text{for} \quad i > j, \\ \hat{p}_{k,l} &= \frac{\lambda_2}{l} \hat{p}_{k,l-1} + \frac{\lambda_3}{l} \hat{p}_{k-1,l-1} + \epsilon_{k,l} \quad \text{for} \quad k \leq l, \end{aligned}$$

where $\epsilon_{i,j}$ and $\epsilon_{k,l}$ are the random errors associated with the linear models. We need to use both models in order to be able to estimate all three parameters (the first model would allow us to estimate only λ_1 and λ_3, and the second model, only λ_2 and λ_3).

To estimate the probability $p_{r,s}$, we use the maximum likelihood estimator $\hat{p}_{r,s} = f_{r,s}/n$. The random variable $f_{r,s}$ has a binomial distribution with parameters $(n, p_{r,s})$, so that

$$\begin{aligned} E(\hat{p}_{r,s}) &= E(f_{r,s}/n) = p_{r,s}, \\ Var(\hat{p}_{r,s}) &= p_{r,s}(1 - p_{r,s})/n; \\ \text{we also have } Cov(\hat{p}_{r,s}, \hat{p}_{r',s'}) &= -p_{r,s} p_{r',s'}/n, \quad \text{for} \quad (r, s) \neq (r', s'). \end{aligned} \tag{3}$$

Propositions 1 and 2 derive some properties of the random errors. Proposition 2 is obtained by rewriting the ϵ's in terms of the $\hat{p}$'s and using the second and third relationship in (3); the proofs of Propositions 1 and 2 can be found in Baba (2005).

Proposition 1: The errors $\epsilon_{i,j}$ and $\epsilon_{k,l}$ are such that

$$
\begin{aligned}
E(\epsilon_{i,j}) &= 0 \quad \text{and} \quad E(\epsilon_{k,l}) = 0, \\
Var(\epsilon_{i,j}) &= \frac{1}{n}\left(p_{i,j} + \left(\frac{\lambda_1}{i}\right)^2 p_{i-1,j} + \left(\frac{\lambda_3}{i}\right)^2 p_{i-1,j-1}\right), \\
Var(\epsilon_{k,l}) &= \frac{1}{n}\left(p_{k,l} + \left(\frac{\lambda_2}{l}\right)^2 p_{k,l-1} + \left(\frac{\lambda_3}{l}\right)^2 p_{k-1,l-1}\right).
\end{aligned}
$$

Proposition 2: The covariance between the errors $\epsilon_{r,s}$ and $\epsilon_{r',s'}$ where $|r - r'| \geq 2$ or $|s - s'| \geq 2$ is equal to zero.

Note that the results in Propositions 1 and 2 are not asymptotic, but are valid for any value of n.

The model can be rewritten as

$$\widehat{Y} = \widehat{X}\theta + \omega,$$

where the vectors $\widehat{Y}$, θ, ω and the matrix $\widehat{X}$ are defined as

$$\widehat{Y}_{c\times 1} = 1/n(f_{0,1}, f_{0,2}, \ldots, f_{0,v}, f_{1,0}, f_{1,1}, \ldots, f_{1,v}, \ldots, f_{u,0}, f_{u,1}, \ldots, f_{u,v})',$$

$$\theta = (\lambda_1, \lambda_2, \lambda_3)',$$

$$\omega_{c\times 1} = (\epsilon_{0,1}, \epsilon_{0,2}, \ldots, \epsilon_{0,v}, \epsilon_{1,0}, \epsilon_{1,1}, \ldots, \epsilon_{1,v}, \ldots, \epsilon_{u,0}, \epsilon_{u,1}, \ldots, \epsilon_{u,v})',$$

$$
\text{and } \widehat{X}_{c\times 3} = \begin{pmatrix}
0 & f_{0,0}/1 & 0 \\
0 & f_{0,1}/2 & 0 \\
\vdots & \vdots & \vdots \\
0 & f_{0,v-1}/v & 0 \\
f_{0,0}/1 & 0 & 0 \\
0 & f_{1,0}/1 & f_{0,0}/1 \\
\vdots & \vdots & \vdots \\
0 & f_{1,v-1}/v & f_{0,v-1}/v \\
\vdots & \vdots & \vdots \\
f_{u-1,0}/u & 0 & 0 \\
f_{u-1,1}/u & 0 & f_{u-1,0}/u \\
\vdots & \vdots & \vdots \\
0 & f_{u,v-1}/v & f_{u-1,v-1}/v
\end{pmatrix},
$$

with $c = (u+1) \times (v+1) - 1$.

This model, linear in its parameter θ, is such that $E(\omega)$=0 and the terms of Σ, the variance-covariance matrix of vector ω, are given in Propositions 1, 2 and in Baba (2005).

Definition: The minimum quadratic distance estimator (QDE) of the parameter θ is the value $\tilde{\theta} = (\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3)'$ which minimizes the distance

$$\omega' \Sigma^{-1} \omega = (\widehat{Y} - \widehat{X}\theta)' \Sigma^{-1} (\widehat{Y} - \widehat{X}\theta), \tag{4}$$

over all θ in the parameter space. This distance will be minimal when θ is equal to

$$\tilde{\theta} = (\widehat{X}' \Sigma^{-1} \widehat{X})^{-1} (\widehat{X}' \Sigma^{-1} \widehat{Y}). \tag{5}$$

This is not an estimator in the usual sense, because the variance-covariance matrix Σ is a function of the unknown parameter θ. To calculate the value of $\tilde{\theta}$, we can try to minimize (4) directly over $(\lambda_1, \lambda_2, \lambda_3)$ or use an iterated reweighted least-squares (IRWLS) algorithm similar to the one used by Huard (1995) for the Katz family.

For this algorithm, we first replace the variance-covariance matrix Σ by the identity matrix I in formula (5) to obtain a first estimator of θ, $\tilde{\theta}_1$. Using this value in Σ, we obtain a first estimate of Σ, $\tilde{\Sigma}_1$, from which we recalculate the estimator of θ from formula (5) and obtain a second estimator $\tilde{\theta}_2$. Repeating this procedure, we obtain a sequence of estimating values of θ, $\{\tilde{\theta}_1, \tilde{\theta}_2, \ldots\}$, and a sequence of estimated variance-covariance matrices $\{\tilde{\Sigma}_1, \tilde{\Sigma}_2, \ldots\}$. To each estimated value $\tilde{\theta}_i$ corresponds an estimated matrix $\tilde{\Sigma}_i$. Huard (1995) shows that the sequence of the values $\{\tilde{\theta}_i\}$ converges to the true parameter value and the sequence of matrices also converges. Luong and Garrido (1993) show that the QDE $\tilde{\theta}$ is asymptotically normally distributed

$$\sqrt{n}(\tilde{\theta} - \theta) \sim N(0, (\widehat{X}' \Sigma^{-1} \widehat{X})^{-1}).$$

3 ASYMPTOTIC EFFICIENCY

The efficiency of an estimator is defined as the ratio of the generalized variance of the maximum likelihood estimator over the generalized variance of the estimator under study, where the generalized variance of an estimator is equal to the determinant of its variance-covariance matrix.

In Table 1, we calculated this asymptotic efficiency for 3 estimation methods investigated by Kocherlakota and Kocherlakota (1992) and Loukas, Kemp and Papageorgiou (1986), the methods of moments (MM), frequency (0,0) (DZ) and even points (EP), compared to the efficiency of the QDE, for various values of $\lambda_1^* = \lambda_2^*$ and the ratio $R = \lambda_3/\lambda_1^*$ between 0 and 1, where $\lambda_1^* = \lambda_1 + \lambda_3$ and $\lambda_2^* = \lambda_2 + \lambda_3$.

For the QDE, since the Jacobian of the transformation from $\theta^* = (\lambda_1^*, \lambda_2^*, \lambda_3)$ to $\theta = (\lambda_1, \lambda_2, \lambda_3)$ is equal to 1, the generalized variances of $\tilde{\theta}$ and $\tilde{\theta}^*$ will be equal. The generalized variance of the QDE is equal to

$$|\Sigma_{QDE}| = \frac{1}{n^3}\det(\widehat{X}'\Sigma^{-1}\widehat{X})^{-1}. \tag{6}$$

For the calculation of the generalized variance of the QDE, we used the largest size possible for Σ in formula (6), which was (63×63) with MATHEMATICA, since we need to take inverses of matrices. The efficiency calculated for the QDE in Table 1 is then a lower bound for the true efficiency.

From Table 1, we see that, for parameter values $\lambda_1^* = \lambda_2^* < 1$, the asymptotic efficiency of the QDE is equal to 1.00. The QDE is still very efficient for $\lambda_1^* = \lambda_2^* = 2$ (efficiency higher than 0.92) and the best estimator. The efficiency of the QDE decreases as $\lambda_1^* = \lambda_2^*$ increases; for $\lambda_1^* = \lambda_2^* = 3$ and the ratio $R = \lambda_3/\lambda_1^* \leq 0.3$, the moment estimator is the best estimator; for $\lambda_1^* = \lambda_2^* = 5$ and $R \leq 0.6$, the moment estimator is the best. For $\lambda_1^* = \lambda_2^* \geq 1$, the double-zero estimator has a low efficiency (less than 30%); it is also less than 30% for the even-points estimator when $\lambda_1^* = \lambda_2^* \geq 4$, and is less than 5% when $R \leq 0.8$.

Bibliography

[1] Baba, A. (2005) *Comparaison de divers estimateurs pour la loi de Poisson bivariée*, Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal.

[2] Huard, L. (1995) *Nouveaux tests d'ajustement pour la loi de Poisson et généralisation*, Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal.

[3] Kocherlakota, S., Kocherlakota, K. (1992) *Bivariate discrete distributions*, Marcel Dekker, Inc., New York.

[4] Loukas, S., Kemp, C.D., Papageorgiou, H. (1986) Even point estimation for the bivariate Poisson distribution, *Biometrika* **73**, 222–223.

[5] Luong, A., Garrido, J. (1993) Minimum quadratic distance estimations for a parametric family of discrete distributions defined recursively, *Australian Journal of Statistic*, **35**, 59–67.

[6] Teicher, H. (1954) On the multivariate Poisson distribution, *Scandinavisk Aktuarietdskrift* **37**, 1–19.

Estimator		$R = \lambda_3/\lambda_1^*$								
	$\lambda_1^* = \lambda_2^*$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
MM	0.05	0.94	0.90	0.86	0.83	0.78	0.73	0.65	0.53	0.35
DZ	0.05	0.71	0.77	0.79	0.80	0.80	0.80	0.77	0.72	0.61
EP	0.05	0.71	0.81	0.87	0.90	0.93	0.95	0.96	0.98	0.99
QDE	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MM	0.10	0.92	0.85	0.80	0.73	0.66	0.58	0.49	0.37	0.21
DZ	0.10	0.61	0.66	0.68	0.68	0.68	0.66	0.63	0.57	0.44
EP	0.10	0.58	0.69	0.76	0.82	0.86	0.90	0.93	0.96	0.98
QDE	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MM	0.50	0.93	0.80	0.66	0.53	0.40	0.30	0.20	0.12	0.05
DZ	0.50	0.40	0.42	0.41	0.39	0.36	0.32	0.27	0.20	0.13
EP	0.50	0.13	0.20	0.27	0.35	0.45	0.55	0.67	0.78	0.89
QDE	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MM	1.00	0.94	0.82	0.67	0.51	0.37	0.24	0.15	0.08	0.03
DZ	1.00	0.26	0.28	0.28	0.27	0.24	0.20	0.16	0.08	0.06
EP	1.00	0.01	0.03	0.05	0.09	0.16	0.26	0.40	0.59	0.79
QDE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MM	2.00	0.95	0.83	0.68	0.51	0.36	0.23	0.13	0.06	0.02
DZ	2.00	0.10	0.11	0.12	0.12	0.12	0.10	0.08	0.05	0.02
EP	2.00	0.00	0.00	0.00	0.00	0.01	0.04	0.10	0.28	0.59
QDE	2.00	0.92	0.93	0.94	0.94	0.94	0.95	0.95	0.96	0.96
MM	3.00	0.95	0.84	0.68	0.52	0.36	0.22	0.13	0.05	0.02
DZ	3.00	0.03	0.04	0.04	0.05	0.05	0.05	0.04	0.03	0.02
EP	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.11	0.42
QDE	3.00	0.60	0.63	0.65	0.67	0.69	0.71	0.74	0.76	0.78
MM	4.00	0.96	0.84	0.69	0.52	0.36	0.23	0.12	0.05	0.01
DZ	4.00	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.00
EP	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.28
QDE	4.00	0.24	0.26	0.28	0.31	0.34	0.37	0.41	0.45	0.50
MM	5.00	0.96	0.84	0.69	0.52	0.36	0.23	0.12	0.05	0.01
DZ	5.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00
EP	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.18
QDE	5.00	0.05	0.07	0.08	0.10	0.11	0.13	0.16	0.20	0.24

Table 1: Asymptotic efficiency of MM, DZ, EP and QDE estimators