ESTIMATING THE PARAMETERS OF THE MULTIVARIATE POISSON DISTRIBUTION USING THE COMPOSITE LIKELIHOOD CONCEPT

Thomas A. Jost, Ramon F. Brcich and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications
Darmstadt University of Technology
Merckstrasse 25, D-64283 Darmstadt, Germany.

ABSTRACT

We address estimation for the multivariate Poisson distribution with second order correlation structure. Existing estimators such as maximum likelihood estimators are too computationally expensive whereas the moment estimator has low efficiency. The proposed estimator uses on the concept of composite likelihood and is, in terms of computational complexity and efficiency, in between a simple moment estimator and the complex maximum likelihood approach.

1. INTRODUCTION

Many problems in signal processing deal with discrete-valued data as for images, categorical data, bio-signal analysis or data networking. While univariate discrete distributions are well known and frequently used, specially the Poisson distribution is popular because of simplicity, multivariate extensions can be rarely seen in the literature. The most common approximations for the multivariate Poisson are the independent model, where the joint density is the product of independent marginal densities, and the multivariate normal approximation. When the mean is low compared to the dimension and significant correlation is present both approximations can be far from optimal.

Recently a paper [1] was published addressing maximum likelihood estimation for multivariate Poisson distributions and [2] for bayesian estimation which allow for second-order correlation. The disadvantage of those algorithms is that the computational complexity makes it difficult to estimate the parameters of high dimensional Poisson distributions e.g. 8. To overcome this problem a simple moment estimator could be used, but it has low efficiency if the correlation between the marginals is non-zero [3]. The proposed estimator is based on the concept of composite likelihood estimation by Lindsay [4], using a pairwise log-likelihood approximation. The method is a compromise between the moment estimator and maximum likelihood estimators in terms of computational complexity and efficiency.

The paper is organized as follows: Section 2 recalls the structure of a multivariate Poisson distribution derived from a multivariate reduction scheme. Section 3 gives the Cramer-Rao bound and both the moment estimator and the composite likelihood estimator, which are compared by simulations in Section 4.

2. THE MULTIVARIATE POISSON DISTRIBUTION

Using the model described in [5] to derive the multivariate Poisson distribution, a vector $\mathbf{X} = (X_1, \ldots, X_m)^T$, where $(\cdot)^T$ denotes transpose, is defined by the linear equation

$$\mathbf{X} = A\mathbf{Y} \tag{1}$$

where $A = [I_m, A_2]$ with I_m being the identity matrix of dimension m and A_2 a $\left[m \times \frac{1}{2}m(m-1)\right]$ matrix composed of all possible unique column vectors having two elements being one and the rest zero. For example, for a trivariate Poisson process A is

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

The vector $\mathbf{Y} = \left(Y_1, \ldots, Y_{m+\frac{1}{2}m(m-1)}\right)^T$ consists of univariate Poisson distributed random elements Y_i with parameter λ_i where Y_i is independent of Y_j for $1 \leq i, j \leq m + \frac{1}{2}m(m-1)$ and $i \neq j$. Because a sum of independent Poisson distributed random variables results in a Poisson distributed random variable the model in eq. (1) describes $\mathbf{X}$ as a multivariate Poisson distributed vector with parameter $\boldsymbol{\lambda} = \left(\lambda_1, \ldots, \lambda_{m+\frac{1}{2}m(m-1)}\right)^T$ and probability mass function

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda}) = \exp\left(-\sum_{k=1}^{m+\frac{1}{2}m(m-1)} \lambda_k\right) \tag{2}$$

$$\times \sum \cdots \sum_{\mathbf{y} \in \mathcal{C}(\mathbf{x})} \left(\prod_{i=1}^{m} \frac{\lambda_i^{x_i - a_i^- \mathbf{y}}}{(x_i - a_i^- \mathbf{y})!}\right) \left(\prod_{l=1}^{\frac{1}{2}m(m-1)} \frac{\lambda_l^{y_{l-m}}}{y_{l-m}!}\right)$$

where a_i^- is the ith row vector of the matrix A_2 and $\mathcal{C}(\mathbf{x})$ is the region of all possible events $\mathbf{y} = (y_1, \ldots, y_{\frac{1}{2}m(m-1)})^T$ for the random variables $Y_{m+1}, \ldots, Y_{m+\frac{1}{2}m(m-1)}$ under a given realization $\mathbf{x}$, i.e.

$$\mathcal{C}(\mathbf{x}) = \bigcap_{i=1}^{m} \left\{a_i^- \mathbf{y} \le x_i\right\} .$$

It should be noted that by the given description of A we restrict ourself to a second-order description between the marginals in $\mathbf{X}$ with only positive correlation.

3. ESTIMATORS

Let $\mathbf{x}(k) = (x_1(k), \ldots, x_m(k))^T$, $k = 1, \ldots, N$, be independent observations of the multivariate Poisson process $\mathbf{X}$.

3.1. Moment Estimator

The moment estimator $\hat{\lambda}_i^M$ for λ_i, $i = m+1, \ldots, m + \frac{1}{2}m(m-1)$, is the sample covariance between the two marginal processes $X_u(n)$ and $X_v(n)$ with $1 \le u, v \le m$ and $u \ne v$ where i is the number of the unique column of A where for rows u and v the elements in A are both 1.
The remaining parameters λ_i for $i = 1, \ldots, m$ are easily estimated using the sample mean $\overline{\mathbf{X}}$ of $\mathbf{X}$ and the equation:

$$\begin{bmatrix} \hat{\lambda}_1^M \\ \vdots \\ \hat{\lambda}_m^M \end{bmatrix} = \overline{\mathbf{X}} - A_2 \begin{bmatrix} \hat{\lambda}_{m+1}^M \\ \vdots \\ \hat{\lambda}_{m+\frac{1}{2}m(m-1)}^M \end{bmatrix} .$$

In this equation, we have to subtract the correlation terms because the sample mean is a biased estimator for $\lambda_1, \ldots, \lambda_m$.

3.2. Composite Likelihood Estimator

The pairwise composite likelihood estimator is defined in [6, 7] where it was used for estimating the parameters of a mixed Poisson distribution and a multivariate survival model. Defining the bivariate marginal log-likelihood function between the random elements X_u, and X_v $l_{uv}(\boldsymbol{\lambda})$ as

$$l_{uv}(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{n=1}^{N} \log f_{X_u X_v}(x_u(n), x_v(n)|\boldsymbol{\lambda}) , \quad (3)$$

where

$$f_{X_u X_v}(x_u(n), x_v(n)|\boldsymbol{\lambda}) = e^{\lambda_u + \lambda_v + a_u^- \boldsymbol{\lambda}_S + a_v^- \boldsymbol{\lambda}_S - \lambda_i} \times \sum_{y_{uv}=0}^{\min(x_u(n), x_v(n))} \frac{(\lambda_u + a_u^- \boldsymbol{\lambda}_S - \lambda_i)^{x_u(n)-y_{uv}}}{(x_u(n) - y_{uv})!} \times \frac{(\lambda_v + a_v^- \boldsymbol{\lambda}_S - \lambda_i)^{x_v(n)-y_{uv}}}{(x_v(n) - y_{uv})!} \cdot \frac{\lambda_i^{y_{uv}}}{y_{uv}!} , \quad (4)$$

using the subset of $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}_S = \left(\lambda_{m+1}, \ldots, \lambda_{m+\frac{1}{2}m(m-1)}\right)^T$, and i is the number of the unique column of A where for rows u and v the elements in A are both 1. The composite log-likelihood function $l(\boldsymbol{\lambda})$ is further defined as the sum of all bivariate log-likelihood functions as

$$l(\boldsymbol{\lambda}) = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{uv} \cdot l_{uv}(\boldsymbol{\lambda}) \quad (5)$$

where w_{uv} is a constant weight for $l_{uv}(\boldsymbol{\lambda})$. For simplicity it is common to set $w_{uv} = 1$ for $1 \le u < v \le m$. The composite score function can be easily derived as the sum of bivariate score functions

$$\frac{\partial l(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{uv} \frac{\partial l_{uv}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} . \quad (6)$$

Similarly, the Hessian matrix can be found as

$$H(\boldsymbol{\lambda}) = \frac{\partial^2 l(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} = \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} w_{uv} \frac{\partial^2 l_{uv}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} .$$

By setting

$$\left. \frac{\partial l(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right|_{\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}^{CL}} = 0 \quad (7)$$

the optimal composite likelihood estimate $\hat{\boldsymbol{\lambda}}^{CL}$ for $\boldsymbol{\lambda}$ can be found. To solve eq. (7) an iterative scheme has to be used.

3.3. Cramer-Rao Bound

The Fisher information matrix $\mathcal{I}(\boldsymbol{\lambda})$ for one sample can be calculated as

$$\mathcal{I}(\boldsymbol{\lambda}) = \mathsf{E}\left[\left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right) \left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right)^T\right]$$

where $\mathsf{E}[\cdot]$ denotes the expected value. The Cramer-Rao bound $\sigma^2_{CRB,\hat{\lambda}_i}$ is [8]

$$\boldsymbol{\sigma}^2_{\boldsymbol{CRB},\hat{\boldsymbol{\lambda}}} = \frac{\operatorname{diag}\left[\mathcal{I}(\boldsymbol{\lambda})^{-1}\right]}{N} = \left(\sigma^2_{CRB,\hat{\lambda}_1}, \ldots, \sigma^2_{CRB,\hat{\lambda}_{m+\frac{1}{2}m(m-1)}}\right)^T \quad (8)$$

where $\operatorname{diag}[B]$ returns the main diagonal of matrix B. By using the recurrence relation for the multivariate Poisson distribution [9]

$$x_i f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda}) = \lambda_i f_{\mathbf{X}}(\mathbf{x} - e_i|\boldsymbol{\lambda}) + \sum_{k=1}^{\frac{1}{2}m(m-1)} a_{ik} \lambda_{m+k} f_{\mathbf{X}}(\mathbf{x} - a_k^{|}|\boldsymbol{\lambda}) ,$$

and the first order derivative of (2) with respect to $\boldsymbol{\lambda}$, the score function can be found as

$$\frac{\partial \log f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \begin{cases} \dfrac{x_i}{\lambda_i} - 1 - \displaystyle\sum_{k=1}^{\frac{1}{2}m(m-1)} a_{ik} \frac{\lambda_{m+k} f_{\mathbf{X}}(\mathbf{x} - a_k^{|}|\boldsymbol{\lambda})}{\lambda_i f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda})} \\ \qquad \text{for } i = 1, \ldots, m \\ \dfrac{f_{\mathbf{X}}(\mathbf{x} - a_{i-m}^{|}|\boldsymbol{\lambda})}{f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda})} - 1 \\ \qquad \text{for } i = m+1, \ldots, m + \frac{1}{2}m(m-1) \end{cases}$$

where e_i is the ith row unit vector, $a_k^{|}$ the kth column vector of the matrix A_2 and a_{ik} its element in row i and column k. It should be noted that for $m = 2$ the composite likelihood estimator is efficient and therefore reaches the Cramer-Rao bound.

3.4. Computational Complexity

The least computationally complex estimator mentioned here is by far the moment estimator. The composite likelihood estimator is more expensive but less than the maximum likelihood estimator. This can be seen by comparing the score functions. The pairwise composite likelihood concept is based on a bivariate probability mass function therefore it requires only a single summation eq. (4) plus two additional to calculate the composite log-likelihood in eq. (5). On the other hand the maximum likelihood estimator uses the multivariate probability mass function in eq. (2) which involves $\frac{1}{2}m(m-1)$ nested summations, making the estimation process computational expensive unless the number of counts in the observation $\mathbf{x}(k)$ are mostly zero.

4. SIMULATIONS

As a global measure of performance we use the sum of the mean squared errors of the parameters

$$\Sigma_{MSE,E} = \sum_{i=1}^{m+\frac{1}{2}m(m-1)} \left(\hat{\lambda}_i^E - \lambda_i\right)^2 \qquad E \in \{M, CL\}$$

to compare the moment and the composite likelihood estimator. We always calculate $\Sigma_{MSE,E}$ from 1000 Monte-Carlo realizations. This is compared to the Cramer-Rao bound as follows

$$\Sigma_{CRB} = \sum_{i=1}^{m+\frac{1}{2}m(m-1)} \sigma^2_{CRB,\hat{\lambda}_i}.$$

$\sigma^2_{CRB,i}$ was evaluated empirically from (8) using 10000 Monte-Carlo realizations.

Data are generated according to the model in eq. (1). To calculate the composite likelihood estimate a large scale Newton method has been used by solving eq. (7) using the function `fmincon` in MATLAB. We used 10^{-3} and the maximum of the sample mean as lower and upper bound for the constraint optimisation. For initialization purposes the moment estimates have been used.

In Fig. (1) the two methods are compared for different values of the dimension m where $\lambda_1 = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.5$ and $N = 500$. As can be seen, the composite likelihood estimator is, in efficiency, between the maximum likelihood and moment estimators except for $m = 2$ where it is efficient in the statistical sense. The measure Σ_{CRB} could be calculated only up to $m = 6$ because of the computational burden.

In Fig. (2), Fig. (3) and Fig. (4) we have displayed $\Sigma_{MSE,E}$ for $m = 4$ and different cases of $\boldsymbol{\lambda}$ versus N. Fig. (2) shows again the case $\lambda_1 = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.5$, Fig. (3) the case where the marginals are highly correlated with $\lambda_1 = \ldots = \lambda_m = 0.2$ and $\lambda_{m+1} = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 2$ while Fig. (4) shows the opposite where the marginals are weakly correlated with $\lambda_1 = \ldots = \lambda_m = 2$ and $\lambda_{m+1} = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.2$.

In Fig. (5) the results are shown again as relative efficiency between the moment estimator and the composite likelihood estimator.

As can be seen the composite likelihood estimator outperforms the moment estimator in every case shown here. The smallest gap appears when the marginals are weakly correlated which is natural because the moment estimator is efficient in the case of independent marginals.

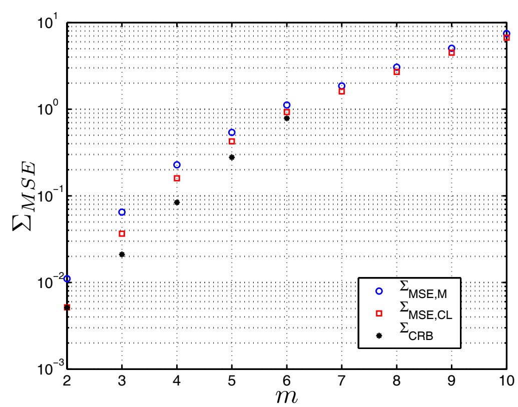

Fig. 1. $\lambda_1 = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.5$ with $N = 500$.

5. CONCLUSION

We have shown that using the concept of composite likelihood to estimate the parameters of a multivariate Poisson

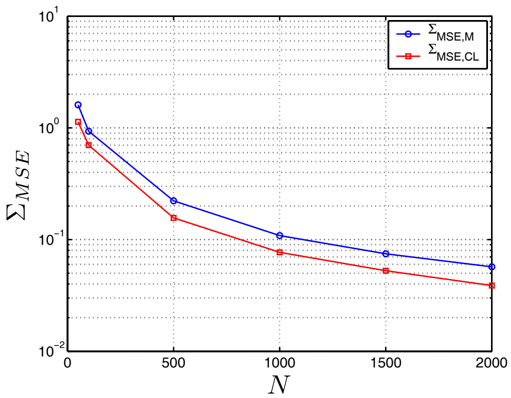

Fig. 2. $\lambda_1 = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.5$ with $m = 4$.

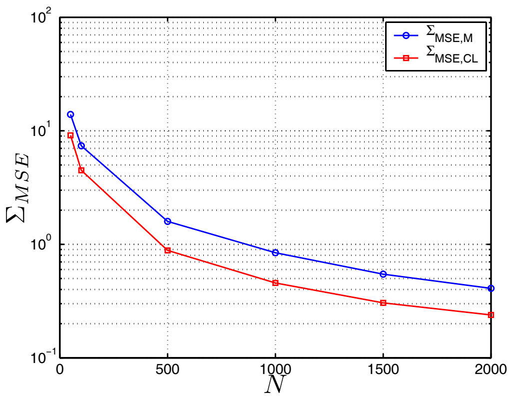

Fig. 3. $\lambda_1 = \ldots = \lambda_m = 0.2$ and $\lambda_{m+1} = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 2$ with $m = 4$.

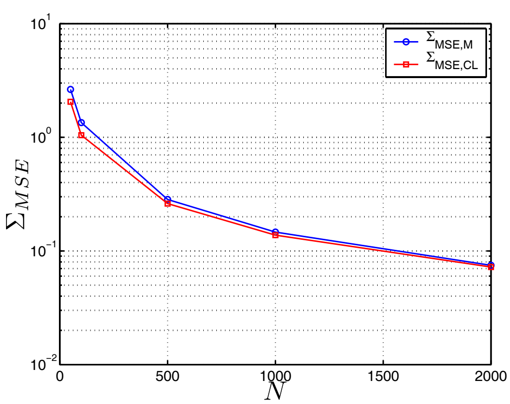

Fig. 4. $\lambda_1 = \ldots = \lambda_m = 2$ and $\lambda_{m+1} = \ldots = \lambda_{m+\frac{1}{2}m(m-1)} = 0.2$ with $m = 4$.

distribution is more efficient than using the moment estimator and is also less computational expensive than using maximum likelihood methods. This makes it preferable in cases where accuracy is needed but maximum likelihood estimation is too complex.

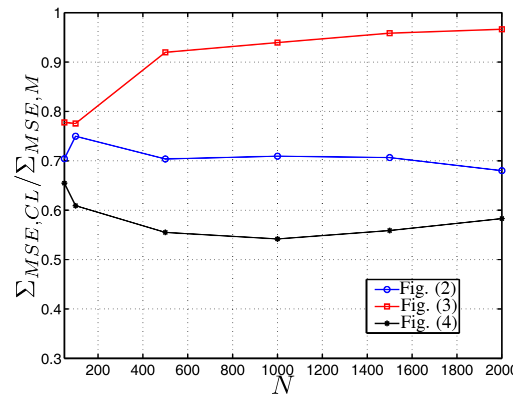

Fig. 5. Relative efficiency between $\Sigma_{MSE,CL}$ and $\Sigma_{MSE,M}$ for the cases of Fig. (2), Fig. (3) and Fig. (4).

Future studies will concentrate on choosing the optimum weights w_{uv} to increase the efficiency of the composite likelihood estimator.

6. ACKNOWLEDGEMENT

The authors would like to thank Dr. D. Karlis of the Departement of Statistics, Athen University of Economics for suggesting corrections and supplying the authors with unpublished papers.

7. REFERENCES

[1] Dimitris Karlis, "An EM algorithm for multivariate Poisson distribution and related models," *Journal of Applied Statistics*, vol. 30, no. 1, pp. 63–77, 2003.

[2] E. Tsionas, "Bayesian analysis of the multivariate Poisson distribution," *Communications in Statistics - Theory and Methods*, vol. 22, pp. 3553–3567, 1999.

[3] P. Holgate, "Estimation for the bivariate Poisson distribution," *Biometrika*, vol. 51, pp. 241–245, 1964.

[4] Bruce G. Lindsay, "Composite likelihood methods," *Contemporary Mathematics*, vol. 80, pp. 221–239, 1988.

[5] Dimitris Karlis and Loukia Meligkotsidou, "Multivariate Poisson regression with covariance structure," *Statistics and Computing*, vol. 15, pp. 255–265, 2005.

[6] Florent Chatelain and Jean-Yves Tourneret, "Composite likelihood estimation for multivariate mixed Poisson distributions," in *SSP 2005*, Bordeaux, France, July 2005.

[7] E.T. Parner, "A composite likelihood approach to multivariate survival data," *Scandinavian Journal of Statistics*, vol. 28, pp. 295–302, 2001.

[8] Steven M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, 1993.

[9] Kazuhiko Kano and Katzutomo Kawamura, "On recurrence relations for the probability function of multivariate generalized Poisson distribution," *Communications in Statistics - Theory and Methods*, vol. 20, no. 1, pp. 165–178, 1991.