COMPOSITE LIKELIHOOD ESTIMATION FOR MULTIVARIATE MIXED POISSON DISTRIBUTIONS

Florent Chatelain and Jean-Yves Tourneret

IRIT/ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
florent.chatelain@tesa.prd.fr, jean-yves.tourneret@tesa.prd.fr,

ABSTRACT

This paper addresses the problem of estimating the parameters of multivariate mixed Poisson distributions. The classical maximum likelihood approach cannot be used for such distributions since they cannot be expressed in simple closed-form. This paper studies an estimation strategy based on the maximization of a so-called composite likelihood criterion. This strategy is compared to a more classical estimator based on the method of moments.

1. INTRODUCTION

Univariate mixed Poisson distributions have received much attention in image processing applications (see for instance [1], [2] and references therein). These applications include active imaging, where the image is obtained from a scene illuminated with laser light [3], or astronomy, where low-flux images are recorded by using photocounting cameras [2]. A univariate mixed Poisson distribution is the distribution of a random variable N such that the conditional distribution of $N|\lambda$ is a Poisson distribution with parameter λ (denoted as $N|\lambda \sim \mathcal{P}(\lambda)$). The parameter λ is also a random variable (called intensity) whose distribution is referred to as structure distribution [1, p. 3] or mixing distribution. When λ has an absolutely continuous distribution defined on $\mathbb{R}^+$ (whose probability density function is denoted as $p(\lambda)$), the probability masses of N can be written:

$$\begin{aligned}\Pr(N=k) &= \int_0^\infty \Pr(N=k|\lambda) f(\lambda) d\lambda, \\ &= \int_0^\infty \frac{\lambda^k}{k!} \exp(-\lambda) f(\lambda) d\lambda. \qquad (1)\end{aligned}$$

Multivariate extensions of mixed Poisson distributions are naturally constructed from a joint intensity distribution $p(\lambda_1, ..., \lambda_d)$ defined on $\mathbb{R}_+^d$. The corresponding masses can be computed as follows:

$$\Pr(\boldsymbol{N}=\boldsymbol{k}) = \int \cdots \int_{\mathbb{R}_+^d} \prod_{\ell=1}^d \frac{(\lambda_\ell)^{k_\ell}}{k_\ell!} \exp(-\lambda_\ell) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \qquad (2)$$

where $\boldsymbol{k} = (k_1, ..., k_d)$ and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_d)$. Some properties of multivariate mixed Poisson distributions (MMPDs) have been recently reported in [4]. In particular, conditions ensuring that MMPDs belong to an exponential family have been derived. This result is important since the maximum likelihood estimator is known to have interesting properties when the observations have a distribution belonging to an exponential family. Unfortunately, conditions on the mixing density ensuring that MMPDs belong to an exponential family are generally too restrictive. This paper studies a composite likelihood approach to estimate the parameters of MMPDs when the mixing distribution is a multivariate Gamma distribution.

This paper is organized as follows. Section 2 recalls some important results on multivariate mixed Poisson distributions. The main properties of the composite likelihood estimator are explained in section 3. Simulation results and conclusions are presented in sections 4 and 5.

2. MMPDS GENERATED BY GAMMA INTENSITIES

2.1. Multivariate Gamma Distributions

A polynomial $P(\boldsymbol{z})$ with respect to $\boldsymbol{z} = (z_1, \ldots, z_d)$ is said to be affine if the one variable polynomial $z_j \mapsto P(\boldsymbol{z})$ can be written $Az_j + B$ (for any $j = 1, \ldots, d$), where A and B are polynomials with respect to the z_i's with $i \neq j$. For any $q \geq 0$ and for any affine polynomial $P(\boldsymbol{z})$, a multivariate Gamma distribution on $\mathbb{R}_+^d$ with shape parameter q and scale parameter $P(\boldsymbol{z})$ (denoted as $\gamma_{q,P}$) is defined by its Laplace transform [5]:

$$\psi_{\gamma_{q,P}}(\boldsymbol{z}) = [P(\boldsymbol{z})]^{-q}, \qquad (3)$$

on an appropriate domain of existence (note that the affine polynomial has to satisfy the condition $P(0) = 1$). De-

This work was supported by the CNRS under MathSTIC Action No. 80/0244.

termining necessary and sufficient conditions on the pair (q, P) such that $\gamma_{q,P}$ exist is a difficult problem. The reader is invited to look at [5] for more details.

This distribution has been used intensively in optics [2] or image processing [3]. Indeed, the complex wave-front amplitude is generally modeled as a zero mean circular Gaussian vector in these applications. Consequently, the vector containing the square modulus of the complex amplitudes is distributed according to multivariate Gamma distribution (with $q = -1$).

2.2. Negative multinomial distributions

For any $q \geq 0$ and for any affine polynomial $P(\boldsymbol{z})$, a negative multinomial distribution $\boldsymbol{N} \sim \mathrm{NM}_{q,P}$ on $\mathbb{N}^d$ is defined by its generating function [6]:

$$E\left(\prod_{k=1}^{d} z_k^{N_k}\right) = [P(\boldsymbol{z})]^{-q}, \tag{4}$$

with the obvious condition $P(1) = 1$. Determining necessary and sufficient conditions on the pair (q, P) such that $\mathrm{NM}_{q,P}$ do exist is a difficult problem (see [6] for more details). Note that, for any affine polynomial P,

$$P_1(z_1, \ldots, z_d) = P(a_1 z_1 + b_1, \ldots, a_d z_d + b_d)$$

is also an affine polynomial, for any real numbers a_i's and b_i's.

2.3. MMPDs generated by multivariate Gamma distributions

The moment generating function of an MMPD $\boldsymbol{N}$ expresses as:

$$\begin{aligned} E\left(\prod_{k=1}^{d} z_k^{N_k}\right) &= E\left(\prod_{k=1}^{d} E(z_k^{N_k}|\lambda_k)\right), \\ &= \psi_\lambda(z_1 - 1, \ldots, z_d - 1), \\ &= \psi_\lambda(\boldsymbol{z} - \boldsymbol{1}), \end{aligned} \tag{5}$$

where $\psi_\lambda(z)$ is the Laplace transform of the intensity distribution. Eq.'s (3,4,5) show that the MMPDs associated to the Gamma distributions $\gamma_{q,P}$ are the **negative multinomial distributions** NM_{q,P_1} with $P_1(\boldsymbol{z}) = P(\boldsymbol{z} - \boldsymbol{1})$. The unknown parameter vector associated to these negative multinomial distributions will be denoted as $\theta = (q, p)$, where q is the shape parameter of the intensity Gamma distribution and p is a vector containing all coefficients of the affine polynomial $P(z)$. As an example, for $d = 2$, the polynomial P can be written as

$$P(z) = 1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2,$$

with $p = (p_1, p_2, p_{12})^T$.

3. COMPOSITE LIKELIHOOD ESTIMATOR

A composite likelihood is a combination of valid likelihood associated to marginal or conditional events. The concept of composite likelihood has been widely studied in the literature (see [7], [8] and references therein) since the seminal paper of Lindsay [9]. Usual composite likelihoods include the marginal likelihood, the pairwise likelihood [9] and the Besag's pseudolikelihood [10]. The composite likelihood estimator is obtained by maximizing the corresponding composite likelihood. The advantage of using composite likelihood instead of standard likelihood is to reduce the computational complexity of the optimization procedure. As a consequence, it allows to handle very complex models, even if the full likelihood cannot be expressed in closed form. This is the case when multivariate mixed Poisson distributions are studied since the joint masses $\Pr(\boldsymbol{N} = \boldsymbol{k})$ cannot be generally computed easily by using (2). This section studies a composite likelihood estimator based on the pairwise likelihood of an MMPD governed by a multivariate Gamma distribution.

3.1. Definition

Consider n time series $\boldsymbol{N} = (N^{(1)}, \ldots, N^{(k)})$, where $N^{(i)}$ is distributed according to a MMPD defined on $\mathbb{R}^d$. The maximum likelihood estimator of the unknown parameters $\theta = (q, p_{ij})$ (where p_{ij} are the coefficients of the polynomial P) requires to optimize the masses of a multinomial distribution defined by its moment generating function (5). This problem is complicated since it is difficult to obtain a tractable expression of the masses $\Pr(\boldsymbol{N} = \boldsymbol{k})$ from (5). As an alternative, we consider the log-likelihood associated with pairwise (N_j^i, N_l^i) (corresponding to the ith time series)

$$l_i^{j,l}(\theta) = \log \Pr(N_j^i = k_j^i, N_l^i = k_l^i).$$

The composite log-likelihood for the ith time series is defined as follows

$$l_i(N^{(i)}; \theta) = \sum_{1 \leq j < k \leq n} w_{j,k} l_i^{j,k}(\theta),$$

where $w_{j,l}$ is an appropriate weight for the pair (N_j, N_l). The composite log-likelihood associated to the n time series Y_n , called *pairwise log-likelihood*, can then be expressed as

$$l(N; \theta) = \sum_{i=1}^{n} l_i(Y^{(i)}; \theta). \tag{6}$$

This paper proposes to estimate the unknown parameter vector θ of an MMPD with moment generating function (5) by maximizing the pairwise log-likelihood (6).
Note that other composite log-likelihood functions (for the ith time series), have been considered in the statistical literature:

- **The Marginal Loglikelihood**

$$l_i^{marg}(N^{(i)};\theta) = \sum_{1\le j\le n} w_i \log \Pr(N_j^i = k_j^i),$$

where w_j is a suitable weight for the jth component N_j^i. This composite log-likelihood is the product of univariate marginal distributions. Consequently, it is generally easy to compute. However, such distribution does not contain any information regarding the covariances of N_j and N_l. Thus, it does not seem interesting for our problem.

- **The Pseudo Loglikelihood**

Another variety of composite log-likelihood, often referred to as *Besag's pseudologlikelihood* is defined by:

$$l_i^{cond}(N^{(i)};\theta) = \sum_{1\le j\le n} w_i \log \Pr(N_j^i = k_j^i | N_{[j]}^i),$$

where $N_{[j]}^i$ denote all the components of N^i except the jth one. The probability $\Pr(N_j^i = k_j^i | N_{[j]}^i)$ provides the distribution of the jth component of $\boldsymbol{N}$ conditioned upon the others components of $\boldsymbol{N}$ (this probability is weighted by w_j). This composite log-likelihood function has shown interesting properties for Markov random fields. However, its expression in the case of MMPDs is more complicated than the pairwise log-likelihood.

As explained above, the choice of a composite likelihood for a practical application is generally motivated by two points: 1) the composite likelihood has to depend on the parameters to be estimated, 2) the computational complexity corresponding to the composite likelihood should be as reduced as possible.

3.2. Properties

The derivative of a composite likelihood is called a composite score function and is denoted by

$$U(N;\theta) = \frac{\partial l(N;\theta)}{\partial\theta}.$$

The composite likelihood estimator is classically obtained by solving the following estimating equations:

$$U(N;\theta) = 0.$$

It is well known that these equations yield a consistent estimator of the unknown parameter vector θ, under appropriate regularity conditions and provided of course that the pairwise log-likelihood depends on θ [11]. Moreover the resulting estimator is asymptotically normal with mean θ and covariance matrix

$$\frac{1}{n} E\left(\frac{\partial U(N;\theta)}{\partial\theta}\right)^{-1} E\left(U(N;\theta)U^T(N;\theta)\right) E\left(\frac{\partial U(N;\theta)}{\partial\theta}\right)^{-T}.$$

These properties result from the structure of the composite score function which is a linear combination of score functions associated with valid log-likelihoods [7, 8, 9, 10].

4. SIMULATION RESULTS

Many simulations have been conducted to validate the previous theoretical results. This section presents some experiments obtained with an intensity vector λ distributed according to a multivariate Gamma distribution. The covariances between the different components of λ have been adjusted as follows:

$$\mathrm{cov}(\lambda_k,\lambda_l) \propto \rho^{|k-l|}, \tag{7}$$

where $\propto$ means proportional to. This covariance structure is well suited to spatial data since the covariance between different observations vanishes when the distance between these observations increases. This parametrization has also the advantage to require only one unknown parameter ρ.

Generation of intensities

The generation of the intensities has been performed as follows:

- simulate q independent multivariate Gaussian vectors of $\mathbb{R}^d$ denoted as $X_1,\ldots,X_q$ with means 0_d and the following $d\times d$ covariance matrix:

$$C = (c_{i,j})_{1\le i,j\le d} = \left(\rho^{\frac{|i-j|}{2}}\right)_{1\le i,j\le d},$$

- compute the kth component of the intensity vector as $\lambda_k = \frac{1}{q}\sum_{1\le i\le q}(X_i^k)^2$, where X_i^k is the kth component of X_i.

It is well known that the random vector $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_d)$ obeys a multivariate Gamma distribution whose marginals are univariate Gamma distributions $\gamma_{1/\xi,1/\xi}$ where $\xi = \frac{2}{q}$. Moreover, the covariance matrix of $\boldsymbol{\lambda}$ can be computed as follows:

$$E(\lambda_k\lambda_l) = \frac{1}{q^2}\sum_{1\le i\le p}\sum_{1\le j\le p} E((X_i^k)^2(X_j^l)^2).$$

The independence between vectors $X_1,\ldots,X_q$ yields

$$E((X_i^k)^2(X_j^l)^2) = E((X_i^k)^2)E((X_j^l)^2) = 1,\ \ \forall\, i\ne j.$$

Moreover,

$$\begin{aligned} E((X_i^k)^2(X_i^l)^2) &= 2E(X_i^kX_i^l)E(X_i^kX_i^l) + E((X_i^k)^2)E((X_i^l)^2),\\ &= 2\rho^{|k-l|}+1, \end{aligned}$$

for $1\le i\le q$. The covariances of (λ_k,λ_l) can be finally expressed as:

$$\mathrm{cov}(\lambda_k,\lambda_l) = \frac{2}{q}\rho^{|k-l|} = \xi\rho^{|k-l|}.$$

This result is in good agreement with (7).

Generation of the MMPDs

It is then possible to generate the MMPD vector $\boldsymbol{N}$ conditionally upon $\boldsymbol{\lambda}$, as $\boldsymbol{N}|\boldsymbol{\lambda} \sim \mathcal{P}(\boldsymbol{\lambda})$. It is important to note that the distribution of $\boldsymbol{N}$ is governed by only two parameters ξ and ρ.

4.1. Marginal Distribution of (N_i, N_j)

The MMPD of $\boldsymbol{N}$ is a negative multinomial distribution (as shown in section 2.2). As a consequence, the marginal distributions of the pairs (N_i, N_j) are negative binomial distributions whose moment generating functions can be written:

$$E\left(z_i^{N_i} z_j^{N_j}\right) = \left[\frac{(1-a)(1-b)-c}{1-az_i-bz_j+(ab-c)z_iz_j}\right]^{-p}. \quad (8)$$

Note that the parameters p, a, b, c (of $\mathbb{R}^+$) have to satisfy the constraint $c \leq (1-a)(1-b)$. By expanding into Taylor series the right hand term of (8), the masses of (N_i, N_j) can be expressed in closed form as:

$$\begin{aligned}\Pr(N_i = m, N_j = n) &= a^m b^n ((1-a)(1-b)-c)^p \\ &\times \sum_{k=0}^{min(m,n)} C^k_{q+k-1} C^{m-k}_{q+m-1} C^{n-k}_{q+n-1} \left(\frac{c}{ab}\right)^k. \quad (9)\end{aligned}$$

The model parameters a, b, c and p can then be expressed as a function of ρ and ξ:

$$\begin{aligned} a = b &= \frac{\xi + \xi^2(1-\rho^{|i-j|})}{1+2\xi+\xi^2(1-\rho^{|i-j|})}, \\ c &= \frac{\xi^2\rho^{|i-j|}}{\left(1+2\xi+\xi^2(1-\rho^{|i-j|})\right)^2}, \quad p = \frac{1}{\xi} \end{aligned}$$

4.2. Parameter Estimation

Eq. (9) shows that the marginal distribution of (N_i, N_j) depends on the model parameters ξ and ρ via a, b, c and p. As a consequence, the composite pairwise log-likelihood (6) seems to be interesting to estimate the parameters of the MMPD associated to $\boldsymbol{N}$. Note again that the properties given in section 3.2 (including consistency and asymptotic normality) are satisfied for MMPDs. The simulations presented in this paper have been obtained with uniform weights $\omega_i = 1, \quad \forall i = 1, ..., n$, so that all pairs (N_i, N_j) uniformly contribute to the composite likelihood. However, other strategies might also also be interesting. For instance, appropriate weights could be chosen in order to mitigate the influence of pairs between non-neighboring observations (which should be less informative in the framework of spatial data). This strategy might reduce the optimization complexity. The optimization procedure used to minimize the negative composite log-likelihood is the direct geometrical Nelder Mead Simplex method (MATLAB function fminsearch.m) . It is important to note that this method does not require costly gradient computations.

In order to appreciate the interest of the proposed composite likelihood method, the unknown parameters ξ and ρ have also been estimated by the classical method of moments. This method is based on the following equations:

$$\begin{aligned} \text{var}(N_i) &= 1 + \xi, && \forall\, 1 \leq i \leq d, \\ \text{cov}(N_i, N_j) &= \rho^{|i-j|}\xi, && \forall\, 1 \leq i \neq j \leq d. \end{aligned}$$

The first equation allows to estimate ξ whereas the parameter ρ can be obtained by using the covariances $\text{cov}(N_i, N_j)$. Note that several methods have been implemented to estimate ρ. Methods based on weighted log-log regressions do not yield better estimation than estimates constructed from the lag-one pairwise. This can be explained by the fact that non-neighboring observations are less informative in our model. As a result, giving too much importance to non-neighboring pairwise leads to bad estimations.

The first simulations show the mean square errors (MSEs) of the estimated parameters ξ and ρ for two different correlation structures ($\rho = 0.5$ and $\rho = 0.8$) as a function of the number of time series n. The number of Monte Carlo runs is 50 for figures 1 and 2 and 500 for figures 3 and 4. The other parameters for this example are $\rho = 0.8$, $q = 8$ (hence $\xi = 1/4$) and $d = 12$. The triangle curve corresponds to the estimator of moments whereas the circle curve corresponds to the composite likelihood estimator. These figures show the interest of the composite likelihood approach, which is much more efficient for this problem than the moment methods, especially for small values of n (number of time series).

Figures 5 and 6 compare the theoretical asymptotic variances of the composite likelihood estimates (provided in section 3.2) with the estimated variances (obtained from 50 Monte Carlo runs). The results are clearly in good agreement.

The exact and estimated distributions of the estimates $\hat{\rho}$ and $\hat{\xi}$ are displayed in figures 7 and 8. The histograms have been obtained from 2000 Monte Carlo runs (the other parameters are $n = 5000$ and $\rho = 0.5$). These figures show that the asymptotic distribution given in section 3.2 is valid for $n = 5000$.

5. CONCLUSIONS

This paper has studied a new strategy for estimating the parameters of multivariate mixed Poisson distributions (MMPDs). The proposed methodology is based on the maximization of an appropriate composite loglikelihood criterion. Simulation results have shown that this method outperforms the classical method of moments for MMPDs. Future works include 1) generalisation of the composite likelihood approach to more sophisticated structures (e.g. by assuming that $X_1, \ldots, X_q$ are autoregressive sequences) and 2) comparison with a weighted non linear least squares approach. The application of the proposed composite likelihood estimation strategy to real images is also under investigation.

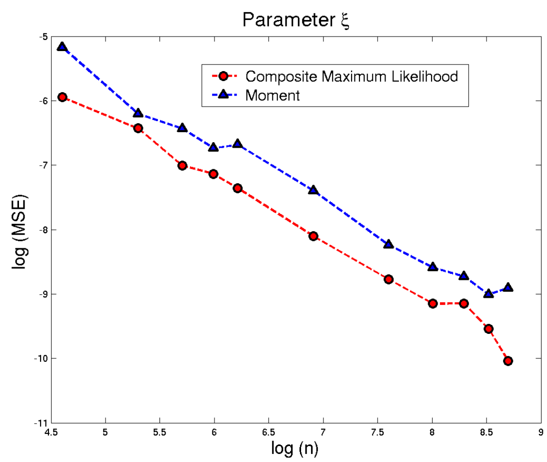

Fig. 1. log MSEs for parameter ξ.

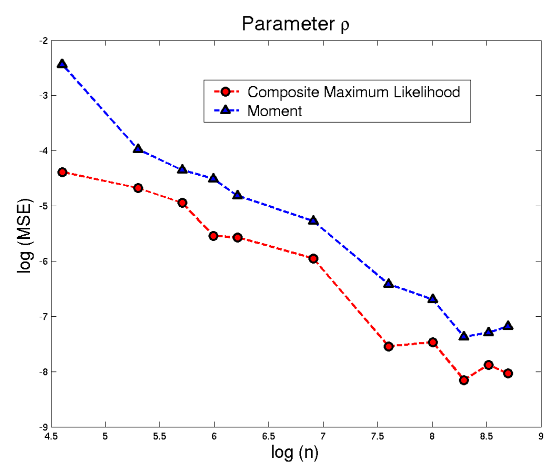

Fig. 2. log MSEs for parameter ρ.

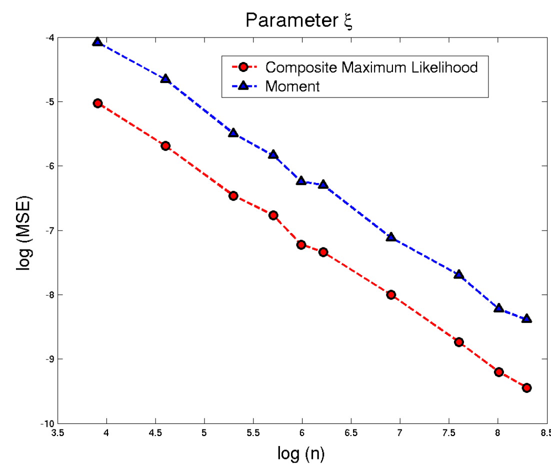

Fig. 3. log MSEs for parameter ξ.

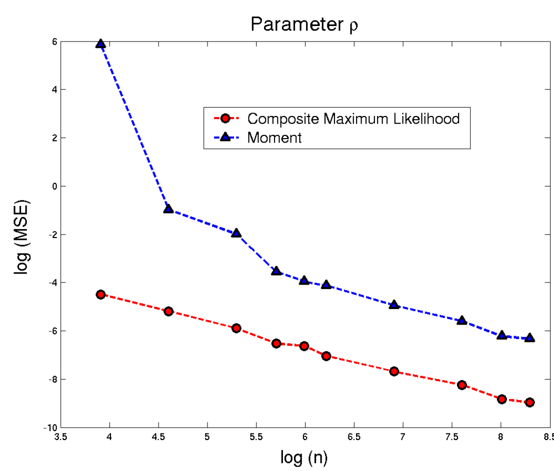

Fig. 4. log MSEs for parameter ρ.

6. REFERENCES

[1] J. Grandell, *Mixed Poisson Processes*. London: Chapman & Hall, 1997.

[2] J. Goodman, *Statistical Optics*. New York: Wiley, 1985.

[3] F. Goudail, N. Roux, and P. Réfrégier, "Performance parameters for detection in low-flux coherent images," *Optical Letters*, vol. 28, pp. 81–83, Jan. 2003.

[4] A. Ferrari, G. Letac, and J.-Y. Tourneret, "Multivariate mixed Poisson distributions," in *EUSIPCO-04* (F. Hlawatsch, G. Matz, M. Rupp, and B. Wistawel, eds.), (Vienna, Austria), pp. 1067–1070, elsevier, Sept. 2004.

[5] P. Bernardoff, "Which multivariate Gamma distributions are infinitely divisible?," *Bernoulli*, to be published.

[6] P. Bernardoff, "Which negative multinomial distributions are infinitely divisible?," *Bernoulli*, vol. 9, pp. 877–893, Dec. 2003.

[7] D. R. Cox and N. Reid, "A note on pseudolikelihood constructed from marginal densities," tech. rep., available at http://www.utstat.toronto.edu/Reid/research.html, 2003.

[8] Y. Zhao and H. Joe, "Composite likelihood estimation in multivariate data analysis," *The Canadian Journal of Statistics*, vol. 33, no. 3, p. to appear, 2005.

[9] B. G. Lindsay, "Composite likelihood methods," *Contemporary Mathematics*, vol. 50, pp. 221–239, 1988.

[10] J. Besag, “Spatial iteraction and the statistical analysis of lattice systems (with discussion),” *J. R. Statist. Soc.*, vol. B36, pp. 192–236, 1974.

[11] V. P. Godambe and C. C. Heyde, “Quasi-likelihood and optimal estimation,” *Int. Statist. Rev.*, vol. 55, pp. 231–234, 1987.

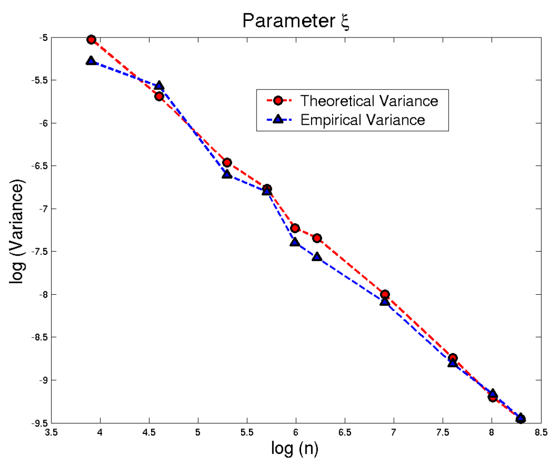

Fig. 5. Asymptotic variance for parameter ξ.

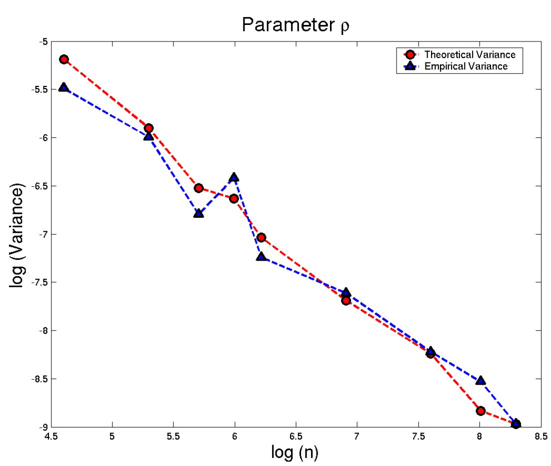

Fig. 6. Asymptotic variance for parameter ρ.

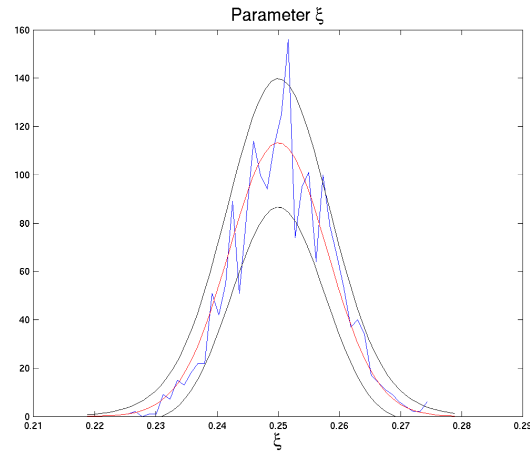

Fig. 7. Estimated and asymptotic distributions of $\hat{\xi}$ with 99% confidence intervals.

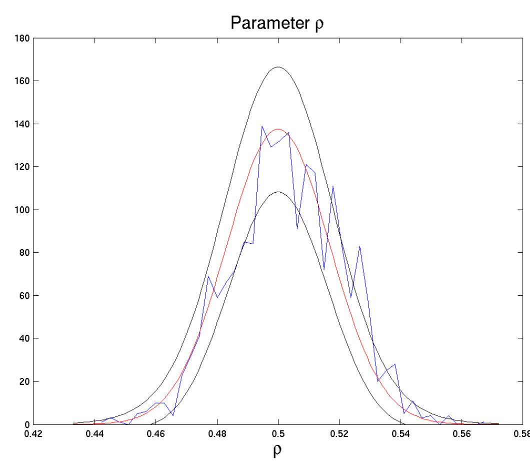

Fig. 8. Estimated and asymptotic distributions of $\hat{\rho}$ with 99% confidence intervals.