PARAMETER ESTIMATION FOR MULTIVARIATE GAMMA DISTRIBUTIONS.

F. Chatelain, J.-Y. Tourneret*, J. Inglada† and A. Ferrari‡*

* IRIT/ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
† CNES, 18 avenue E. Belin, BPI 1219, 31401 Toulouse cedex 9, France
‡ LUAN, Université de Nice-Sophia-Antipolis, 06108 Nice cedex 2, France
{florent.chatelain,jean-yves.tourneret}@enseeiht.fr, jordi.inglada@cnes.fr, ferrari@unice.fr

ABSTRACT

This paper addresses the problem of estimating the parameters of bivariate gamma distributions. The estimators based on the maximum likelihood method and the method of moments are derived for this problem. The asymptotic performance of both estimators are also studied.

1. INTRODUCTION

The univariate gamma distribution is uniquely defined in many statistical textbooks. However, extensions defining multivariate gamma distributions (MGDs) are more controversial. For instance, a full chapter of [1] is devoted to this problem (see also references therein). Most journal authors assume that a vector $\mathbf{X} = (X_1, \dots, X_d)$ is distributed according to an MGD if the marginal distributions of X_i are univariate gamma distributions. However, the family of distributions satisfying this condition is very large. In order to reduce the size of the family of MGDs, S. Bar Lev and P. Bernardoff recently defined MGDs by the form of their moment generating function (or Laplace transform) [2], [3]. The main contribution of this paper is to study estimators for the parameters of bivariate gamma distributions (BGDs) defined in [2], [3]. These distributions are interesting for image registration as discussed below.

Given two remote sensing images of the same scene I, the reference, and J, the secondary image, the registration problem can be defined as follows: determine a geometric transformation T which maximizes the correlation coefficient between image I and the result of the transformation $T \circ J$. A fine modeling of the geometric deformation is required for the estimation of the coordinates of every pixel of the reference image inside the secondary image. The geometric deformation is modeled by local rigid displacements [4].

The key element of the image registration problem, is the estimation of the correlation coefficient between the images. This is usually done with an estimation window in the neighborhood of each pixel. In order to estimate the local rigid displacements with a good geometric resolution one needs the smallest estimation window. However, this leads to estimations which may not be robust enough. In order to perform high quality estimations with a small number of samples, we propose to introduce a priori knowledge about the image statistics. In the case of power radar images, it is well known that the pixels follow a gamma distribution [5]. Therefore, MGDs seem good candidates for the robust estimation of the correlation coefficient between radar images.

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

This paper is organized as follows. Section 2 recalls some important results on MGDs. Section 3 studies two estimators of the unknown parameters of a BGD. These estimators are based on the classical maximum likelihood method and method of moments. Simulation results illustrating the performance of both estimators are presented in Section 4. Conclusions are finally reported in Section 5.

2. MULTIVARIATE GAMMA DISTRIBUTIONS

2.1 Definitions

A polynomial $P(\mathbf{z})$ with respect to $\mathbf{z} = (z_1, \dots, z_d)$ is *affine* if the one variable polynomial $z_j \mapsto P(\mathbf{z})$ can be written $Az_j + B$ (for any $j = 1, \dots, d$), where A and B are polynomials with respect to the z_i's with $i \neq j$. A random vector $\mathbf{X} = (X_1, \dots, X_d)$ is distributed according to an MGD on $\mathbb{R}_+^d$ with shape parameter q and scale parameter P (denoted as $X \sim \Gamma(q, P)$) if its moment generating function or Laplace transform is defined as follows [3]:

$$\psi_{\gamma_{q,P}}(\mathbf{z}) = \mathsf{E}\left(e^{-\sum_{i=1}^{d} X_i z_i}\right) = [P(\mathbf{z})]^{-q}, \qquad (1)$$

where $q \geq 0$ and P is an affine polynomial. It is important to note the following points:

- the affine polynomial P has to satisfy appropriate conditions including $P(0) = 1$. In the general case, determining 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 [3] for more details,
- by setting $z_j = 0$ for $j \neq i$ in (1), we obtain the Laplace transform of X_i, which is clearly a gamma distribution with shape parameter q and scale parameter p_i, where p_i is the coefficient of z_i in P.

A BGD corresponds to the particular case $d = 2$ and is defined by its Laplace transform

$$\psi(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^{-q}, \qquad (2)$$

with the following conditions

$$p_1 > 0,\ p_2 > 0,\ p_1 p_2 - p_{12} > 0. \qquad (3)$$

In the bi-dimensional case, the conditions (3) ensure that (2) is the Laplace transform of a probability distribution defined on $[0, \infty[^2$. Note again that (2) implies that the marginal distributions of X_1 and X_2 are gamma distributions, i.e. $X_1 \sim \Gamma(q, p_1)$ and $X_2 \sim \Gamma(q, p_2)$.

2.2 Bivariate Gamma pdf

Obtaining tractable expressions for the probability density function (pdf) of a MGD defined by (1) is a challenging problem. However, in the bivariate case, the problem is much simpler. Straightforward computations allow to obtain the following density (see [1, p. 436] for a similar result)

$$f_{2D}(\mathbf{x}) = \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1} x_2^{q-1}}{p_{12}^q \Gamma(q)} f_q(c x_1 x_2) \mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}),$$

where $\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x})$ is the indicator function defined on $[0,\infty[^2$ ($\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}) = 1$ if $x_1 > 0, x_2 > 0$, $\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}) = 0$ else), $c = \frac{p_1 p_2 - p_{12}}{p_{12}^2}$ and $f_q(z)$ is defined as follows

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!\Gamma(q+k)}.$$

Note that $f_q(z)$ is related to the confluent hypergeometric function (see [1, p. 462]).

2.3 BGD Moments

The Taylor series expansion of the Laplace transform ψ can be written:

$$\psi(z_1, z_2) = \sum_{k,l \geq 0} \frac{(-1)^{k+l}}{k!l!} E[X_1^k X_2^l] z_1^k z_2^l.$$

By derivating this expression, the moments of a BGD can be obtained. For instance, the mean and variance of X_i (denoted $E[X_i]$ and $\mathrm{var}(X_i)$ respectively) can be expressed as follows

$$E[X_i] = q p_i, \ \mathrm{var}(X_i) = q p_i^2,$$

for $i = 1, 2$. Similarly, the covariance $\mathrm{cov}(X_1, X_2)$ and correlation coefficient $r(X_1, X_2)$ of a BGD can be easily computed:

$$\begin{aligned} \mathrm{cov}(X_1, X_2) &= E[X_1 X_2] - E[X_1]E[X_2] = q(p_1 p_2 - p_{12}), \\ r(X_1, X_2) &= \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)}\sqrt{\mathrm{var}(X_2)}} = \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \end{aligned}$$

It is important to note that for a known value of q, a BGD is fully characterized by $\theta = (E[X_1], E[X_2], r(X_1, X_2))$ (since θ and (p_1, p_2, p_{12}) are related by a one-to-one transformation). Note also that the conditions (3) ensure that the covariance and correlation coefficient of the couple (X_1, X_2) are both positive.

3. PARAMETER ESTIMATION

The following notations are used in the rest of the paper

$$m_1 = E[X_1], \ m_2 = E[X_2], \ r = r(X_1, X_2),$$

inducing $\theta = (m_1, m_2, r)$. This section addresses the problem of estimating the unknown parameter vector θ from n vectors $\mathbf{X} = (\mathbf{X}^1, \ldots, \mathbf{X}^n)$, where $\mathbf{X}^i = (X_1^i, X_2^i)$ is distributed according to a BGD with parameter vector θ. Note that the parameter q is assumed to be known here, as in most practical applications. However, this assumption could be relaxed.

3.1 Maximum Likelihood Method

3.1.1 Principles

The maximum likelihood (ML) method can be applied in the bivariate case ($d = 2$) since a closed-form expression of the density is available. In this particular case, after removing the terms which do not depend on θ, the log-likelihood function can be written

$$\begin{aligned} l(\mathbf{X};\theta) = &-nq\log(m_1 m_2) - \frac{nq\overline{X}_1}{m_1(1-r)} - \frac{nq\overline{X}_2}{m_2(1-r)} \\ &-nq\log(1-r) + \sum_{i=1}^{n} \log f_q(c X_1^i X_2^i), \end{aligned} \quad (4)$$

where $c = \frac{rq^2}{m_1 m_2 (1-r)^2}$, and $\overline{X}_1 = \frac{1}{n}\sum_{i=1}^n X_1^i$, $\overline{X}_2 = \frac{1}{n}\sum_{i=1}^n X_2^i$ are the sample means of X_1 and X_2. By differentiating the log-likelihood with respect to θ and by noting that $f_q'(z) = f_{q+1}(z)$, the following set of equations is obtained

$$\begin{aligned} &\frac{nq\overline{X}_1}{1-r} - nqm_1 - \frac{r}{(1-r)^2}\frac{q^2}{m_2}\Delta = 0, \\ &\frac{nq\overline{X}_2}{1-r} - nqm_2 - \frac{r}{(1-r)^2}\frac{q^2}{m_1}\Delta = 0, \\ &\frac{nq\overline{X}_1}{(1-r)m_1} + \frac{nq\overline{X}_2}{(1-r)m_2} - nq - \frac{1+r}{(1-r)^2}\frac{q^2}{m_1 m_2}\Delta = 0, \end{aligned}$$

where

$$\Delta = \left(\sum_{i=1}^n X_1^i X_2^i \frac{f_{q+1}(cX_1^i X_2^i)}{f_q(cX_1^i X_2^i)}\right).$$

The maximum likelihood estimators (MLEs) of m_1 and m_2 are then easily obtained

$$\widehat{m_1}_{\mathrm{ML}} = \overline{X}_1, \quad \widehat{m_2}_{\mathrm{ML}} = \overline{X}_2.$$

The MLE of r is obtained by computing the root $r \in]0,1[$ of

$$g(r) = r - 1 + \frac{q}{n\overline{X}_1\overline{X}_2}\left(\sum_{i=1}^n X_1^i X_2^i \frac{f_{q+1}(\widehat{c}X_1^i X_2^i)}{f_q(\widehat{c}X_1^i X_2^i)}\right) = 0, \quad (5)$$

where

$$\widehat{c} = \frac{r}{(1-r)^2}\frac{q^2}{\overline{X}_1\overline{X}_2}.$$

This is achieved by using a Newton-Raphson procedure initialized by the standard empirical correlation coefficient defined in (9). It is possible to show that the function (5) has a unique root in $[0,1[$ provided $\widehat{r}_{\mathrm{Mo}}$ defined in (9) belongs in $[0,1[$. The convergence of the Newton-Raphson procedure is practically obtained after few iterations.

3.1.2 Performance

The asymptotic properties of the ML estimators $\widehat{m_1}_{\mathrm{ML}}$ and $\widehat{m_1}_{\mathrm{ML}}$ can be derived from the univariate gamma distributions $\Gamma(q, p_1)$ and $\Gamma(q, p_2)$. These estimators are obviously unbiased, convergent and efficient. However, the performance of $\widehat{r}_{\mathrm{ML}}$ is more difficult to derive. Of course, the MLE is known to be asymptotically unbiased and asymptotically efficient, under mild regularity conditions. Thus, the mean square error of the estimates can be approximated for large data records by the Cramer-Rao lower bound (CRLB). For unbiased estimators, the CRLB is obtained by inverting the Fisher information matrix. The computation of this matrix requires to determine the negative expectations of second-order derivatives (with respect to m_1, m_2 and r) of $l(\mathbf{X};\theta)$ in (4). Closed-form expressions for the expectations are difficult to obtain because of the term $\log f_q$. In such situation, it is very usual to approximate the expectations by using Monte Carlo methods. This will provide interesting approximations of the ML mean square errors (MSEs) (see simulation results of section 4).

3.2 Method of Moments

3.2.1 *Principles*

This section briefly recalls the principle of the method of moments. Consider a function $\mathbf{h}(.) : \mathbb{R}^M \to \mathbb{R}^L$ and the statistic s_n of size L defined as:

$$\mathbf{s}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{h}(\mathbf{X}^i), \tag{6}$$

where $\mathbf{h}(.)$ is usually chosen such that $\mathbf{s}_n$ is composed of empirical moments. Denote as:

$$\mathbf{f}(\theta) = \mathsf{E}[\mathbf{s}_n] = \mathsf{E}[\mathbf{h}(\mathbf{X}^1)].$$

The moment estimator of θ is constructed as follows:

$$\hat{\theta}_{\mathrm{Mo}} = \mathbf{g}(\mathbf{s}_n),$$

where $\mathbf{g}(\mathbf{f}(\theta)) = \theta$. By considering the function

$$\mathbf{h}(\mathbf{X}) = (X_1, X_2, X_1^2, X_2^2, X_1X_2),$$

the following result is obtained

$$\mathbf{f}(\theta) = [m_1, m_2, m_1^2(1+q^{-1}), m_2^2(1+q^{-1}), m_1m_2(1+rq^{-1})].$$

The unknown parameters (m_1, m_2, r) can then be expressed as functions of $\mathbf{f}(\theta) = (f_1, f_2, f_3, f_4, f_5)$. For instance, the following relations are obtained

$$m_1 = f_1,\ m_2 = f_2,\ r = \frac{f_5 - f_1 f_2}{\sqrt{(f_3 - f_1^2)(f_4 - f_2^2)}}, \tag{7}$$

yielding the standard estimators:

$$\widehat{m_1}_{\mathrm{Mo}} = \overline{X}_1, \quad \widehat{m_2}_{\mathrm{Mo}} = \overline{X}_2, \tag{8}$$

$$\widehat{r}_{\mathrm{Mo}} = \frac{\sum_{i=1}^{n}(X_1^i - \overline{X}_1)(X_2^i - \overline{X}_2)}{\sqrt{\sum_{i=1}^{n}(X_1^i - \overline{X}_1)^2}\sqrt{\sum_{i=1}^{n}(X_2^i - \overline{X}_2)^2}}. \tag{9}$$

3.2.2 *Performance*

The asymptotic performance of the estimator $\hat{\theta}_{\mathrm{Mo}}$ can be derived by imitating the results of [6] derived in the context of time series analysis. A key point of these proofs is the assumption $\mathbf{s}_n \overset{a.s.}{\to} \mathbf{s} = \mathbf{f}(\theta)$ which is verified herein by applying the strong law of large numbers to (6). As a result, the asymptotic mean square error of $\hat{\theta}_{\mathrm{Mo}}$ can be derived:

$$\lim_{n\to\infty} n\mathsf{E}[(\hat{\theta}_{\mathrm{Mo}} - \theta)^2] = \mathbf{G}(\theta)\boldsymbol{\Sigma}(\theta)\mathbf{G}(\theta)^t, \tag{10}$$

where $\mathbf{G}(\theta)$ is the Jacobian matrix of the vector $\mathbf{g}(\cdot)$ at point $\mathbf{s} = \mathbf{f}(\theta)$ and

$$\Sigma(\theta) = \lim_{n\to\infty} nE[(s_n - s)(s_n - s)^T].$$

In the previous example, according to (7), $g : \mathbb{R}^5 \to \mathbb{R}^3$ is defined as follows

$$\mathbf{g}(\mathbf{x}) = \left(x_1, x_2, \frac{x_5 - x_1x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}}\right).$$

The partial derivatives of g_1 and g_2 with respect to $x_i, i = 1,\ldots,5$ are trivial. By denoting $\gamma = \sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}$, those of g_3 can be expressed as

$$\begin{aligned}
\frac{\partial g_3}{\partial x_1} &= -\frac{x_2}{\gamma} + \frac{x_1(x_4 - x_2^2)(x_5 - x_1x_2)}{\gamma^3},\\
\frac{\partial g_3}{\partial x_2} &= -\frac{x_1}{\gamma} + \frac{x_2(x_3 - x_1^2)(x_5 - x_1x_2)}{\gamma^3},\\
\frac{\partial g_3}{\partial x_3} &= \frac{(x_1x_2 - x_5)(x_4 - x_2^2)}{2\gamma^3},\\
\frac{\partial g_3}{\partial x_4} &= \frac{(x_1x_2 - x_5)(x_3 - x_1^2)}{2\gamma^3},\\
\frac{\partial g_3}{\partial x_5} &= \frac{1}{\gamma}.
\end{aligned}$$

The elements of $\Sigma(\theta)$ can be computed from the moments of $\mathbf{h}(\mathbf{X})$ which are obtained by derivating the Laplace transform (2). This allows to compute the asymptotic MSE (10) thanks to the general formula valid for any integers m, n:

$$E[X^mY^n] = m_1^m m_2^n \frac{(q)_m}{q^m}\frac{(q)_n}{q^n}\sum_{k=0}^{\min(m,n)}\frac{(-m)_k(-n)_k}{(q)_k}\frac{r^k}{k!}, \tag{11}$$

where $(a)_k$ is the Pochhammer symbol defined by $(a)_0 = 1$ and

$$(a)_{k+1} = (a+k)\,(a)_k = a(a+1)\ldots(a+k),$$

for any integer k (see [7, p. 256]).

4. SIMULATION RESULTS

Many simulations have been conducted to validate the previous theoretical results. This section presents some experiments obtained with a vector $\mathbf{X} = (X_1, X_2)$ distributed according to a BGD whose Laplace transform is (2).

4.1 Generation

The generation of $\mathbf{X}$ has been performed as follows:

- simulate $2q$ independent multivariate Gaussian vectors of $\mathbb{R}^2$ denoted as $Z^1, \ldots, Z^{2q}$ with means $(0,0)$ and the following 2×2 covariance matrix:

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

- compute the kth component of $\mathbf{X} = (X_1, X_2)$ as $X_k = \frac{m_k}{2q}\sum_{1\le i\le 2q}(Z_k^i)^2$, where Z_k^i is the kth component of Z^i.

By computing the Laplace transform of $\mathbf{X}$, it can be shown that the two previous steps allow to generate random vectors $\mathbf{X} = (X_1, X_2)$ distributed according to a BGD. The marginal distributions of X_1 and X_2 are univariate gamma distributions $\Gamma(q, m_1/q)$ and $\Gamma(q, m_2/q)$. Moreover, the covariance of $\mathbf{X}$ can be computed as follows:

$$E(X_1X_2) = \frac{m_1 m_2}{4q^2} \sum_{i=1}^{2q} \sum_{j=1}^{2q} E[(Z_1^i)^2(Z_2^j)^2].$$

The independence between vectors $Z^1, \ldots, Z^{2q}$ yields

$$E[(Z_1^i)^2(Z_2^j)^2] = E[(Z_1^i)^2]E[(Z_2^j)^2] = 1, \ \forall\, i \neq j.$$

Moreover,

$$\begin{aligned} E[(Z_1^i)^2(Z_2^i)^2] &= 2E(Z_1^i Z_2^i)E(Z_1^i Z_2^i) + E[(Z_1^i)^2]E[(Z_2^i)^2], \\ &= 2r + 1, \end{aligned}$$

for $1 \leq i \leq 2q$. The covariance of (X_1, X_2) and the corresponding correlation coefficient can be finally expressed as:

$$\mathrm{cov}(X_1, X_2) = \frac{m_1 m_2 r}{q}, \quad \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} = r.$$

4.2 Estimation Performance

The first simulations compare the performance of the method of moments with the ML method as a function of n. Note that the possible values of n corresponds to the numbers of pixels of squared windows of size $(2p+1) \times (2p+1)$, where $p \in \mathbb{N}$. These values are appropriate to the image registration problem. The number of Monte Carlo runs is 1000 for all figures presented in this section. The other parameters for this example are $m_1 = 400$, $m_2 = 800$ and $q = 1$. Figures 1 and 2 show the mean square errors (MSEs) of the estimated correlation coefficient (obtained from 1000 Monte Carlo runs) for two different correlation structures ($r = 0.2$ and $r = 0.8$). The circle curves correspond to the estimator of moments whereas the triangle curves correspond to the MLE. These figures show the interest of the ML method, which is much more efficient for this problem than the method of moments. The figures also show that the difference between the two methods is more significative for large values of the correlation coefficient r.

The theoretical asymptotic MSEs of the ML and moment estimators are also depicted on Figs. 1 and 2 (continuous lines). The theoretical MSEs are clearly in good agreement with the estimated MSEs, even for small values of n. This is particularly true for large values of r. Finally, these figures show that "reliable" estimates of r can be obtained for values of n larger than 9×9.

5. CONCLUSIONS

This paper studied maximum likelihood and moment estimators for the parameters of bivariate gamma distributions. The asymptotic performance of these estimators was also investigated. The application of these results to image registration and to change detection is currently under investigation.

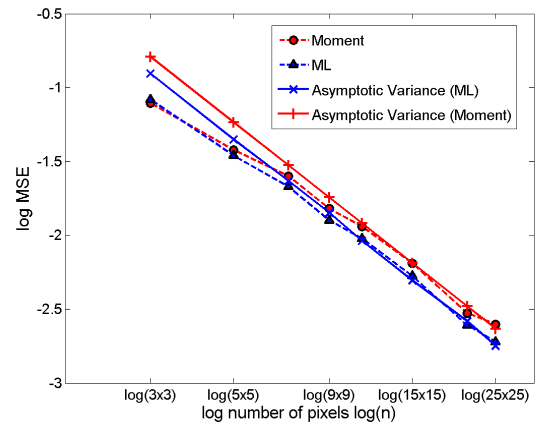

Figure 1: log MSEs versus $\log(n)$ for parameter r ($r = 0.2$).

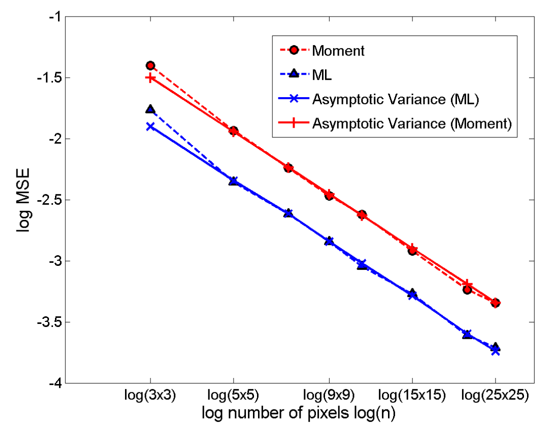

Figure 2: log MSEs versus $\log(n)$ for parameter r ($r = 0.8$).

6. ACKNOWLEDGMENTS

The authors would like to thank G. Letac for fruitful discussions regarding multivariate gamma distributions.

REFERENCES

[1] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, vol. 1. New York: Wiley, 2nd ed., 2000.

[2] S. B. Lev, D. Bshouty, P. Enis, G. Letac, I. L. Lu, and D. Richards, "The diagonal natural exponential families and their classification," *J. of Theoret. Prob.*, vol. 7, pp. 883–928, 1994.

[3] P. Bernardoff, "Which multivariate Gamma distributions are infinitely divisible?," *Bernoulli*, 2006.

[4] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2104–2120, Oct. 2004.

[5] T. F. Bush and F. T. Ulaby, "Fading characteristics of panchromatic radar backscatter from selected agricultural targets," *NASA STI/Recon Technical Report N*, vol. 75, pp. 18460–+, Dec. 1973.

[6] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *International Journal of adaptive control and signal processing*, vol. 3, pp. 191–229, 1989.

[7] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1972.