Maximum Likelihood Estimation of K-distribution Parameters Using Number Theoretic Methods

Da-Peng Li[+], Ying-Qin Sun and Di Yao

Beijing Institute of Technology, School of Information and Electronics

Beijing, China

Abstract—The K-distribution is widely applied in synthetic aperture radar (SAR) image processing. However, the multi-peak complicated likelihood function causes much trouble to obtain the maximum likelihood estimation of K-distribution parameters. Based on the number-theoretic net (NT-net), the computable steps of sequential number-theoretic method for optimization (SNTO) were proposed to get the MLE of the parameters of K-distribution. Comparing with the non-ML estimator Y0.1, we do Monte Carlo trails with different values of shape parameter and different sample sizes. The simulation results show that the proposed method outperforms the fractional moment based technique.

Keywords- K-distribution; MLE; Number-Theoretic Method; SNTO

1. Introduction

With the development of radar technology, the resolution of radar is becoming higher dramatically. It makes the amplitude of radar echo signal biasing the Rayleigh distribution and has a long tail. Thus the classical Rayleigh distribution model is often not suitable for describing the clutter statistics in the case of fine resolution and/or heterogeneous regions. In 1976, Jakeman and Pusey [1] introduced the K-distribution, originally for sea clutter, which since has found wide use also to describe the distribution of land clutter. Many studies [2][3][4] proved that the K-distribution is one of the most successful models for backscattering statistics in synthetic aperture radar (SAR) images.

In the situation of K-distributed clutter, accurate and reliable estimation of the parameters is of great importance for radar detection and estimation problems, such as SAR image segmentation and parametric CFAR detection. Study [5] shows that the performance of CFAR will be degenerate seriously when the error of estimation of K-distribution parameters becomes large. Therefore, many estimators have been discussed recent years. In previous studies, the most widely used methods are the method of moments (MOM). Raghavan [6] has presented a method for estimating K-distribution parameters based on arithmetic and geometric means of the data. Oliver [7] has proposed three famous estimators, namely: contrast (V), normalized log (U) and variance of log (W); Blacknell and Tough [8] have exploited the limit property of estimator Y and proposed an estimator X.

The common characters of these methods are easy to implement, but none of them are the optimum one comparing to MLE. In order to improve the accuracy of the estimator, Wachowiak M P [9] used neural networks to evaluate the parameters of K-distribution. It is complicated and needs training for a period of time, although it has high accuracy and is robust to noise. Theoretically, the ML method provides the smallest errors that approach the Cramer Rao lower bound, however, the ML solutions cannot be derived analytically, and

---

[+] Corresponding author.
 E-mail address: insanegtp99@yahoo.com.cn

numerical search is often needed, which is resource and time consuming [10][11] In this paper, a new ML estimation scheme based on number-theoretic methods will be considered. This scheme needs relative little computation, and can be calculated in parallel which is suitable for practical application. A comparison of errors between this scheme and the MME will be made in different sample sizes for a range of shape parameter values.

2. Problems in MLE of K-distribution

The probability density function (PDF) of K-distribution is given as,

$$f(x) = \frac{2}{a\Gamma(v)}(\frac{x}{2a})^v K_{v-1}(\frac{x}{a}), x \geq 0, v > 0, a > 0, \tag{1}$$

where x is the amplitude of clutter, v is the shape parameter which controls the shape of K-distribution. When v is increasing, the kurtosis of K-distribution goes high. a is the scale parameter which has relationship with the intensity of clutter, and K_{v-1} is the modified Bessel function of the second kind.

Given N independent samples $x=(x_1, x_2, \cdots, x_N)$ which is following the K-distribution, the likelihood function is the product of the marginal PDFs. The log-likelihood function is

$$L_N(\mathbf{x}; a, v) = \ln[\prod_{i=1}^{N} f(x_i; a, v)]$$

$$= N(1-v)\ln(2) - N(1+v)\ln(a) - N\ln(\Gamma(v)). \tag{2}$$

$$+ v\sum_{i=1}^{N}\ln(x_i) + \sum_{i=1}^{N}\ln(K_{v-1}(\frac{x_i}{a}))$$

The maximum likelihood estimates a, v is given by maximizing the log-likelihood function (3).

$$\begin{cases} \frac{\partial}{\partial v}[\sum_{i=1}^{N}\ln(L_N(x; a, v))]\big|_{v=\hat{v}} = 0 \\ \frac{\partial}{\partial a}[\sum_{i=1}^{N}\ln(L_N(x; a, v))]\big|_{a=\hat{a}} = 0 \end{cases} \tag{3}$$

For K-distribution, it's obvious the variable v is in the order of the modified Bessel function, the derivative with respect to it is not a closed form, and thus, Ian R. Joughin [12] derived a ML solution for one parameter of the K-distribution using the generalized Bessel function K-distribution. William J. J. Roberts [13] proposed an ML estimator using the expectation maximization (EM) approach. But all these methods require relative large amounts of computation, and are too expensive for the applications of interest.

3. MLE using number theoretic methods

3.1. Number Theoretic Methods

In parameter estimation theory, MLE is considered as the optimum one and applied widely in many fields. However, there are no common methods to solve a variety of likelihood function in theory. In fact, we have to seek help from numerical calculations, such as Gauss-Newton method, iteration method, simplex method, conjugate gradient method, etc. For the purpose of finding out the global extremum, the likelihood function needs to be single peak and differentiable. But the fact is there are usually many peaks existed in the practical likelihood functions, and sometimes the partial differentiation doesn't exist. In such situation it's impossible to solve the problem using these numerical methods. Although some methods developed to find the global extremum in multi-peak functions, such as Monte Carlo method, Simulated Annealing, the rates of convergence are still very slow for application. In 1978, Hua, L.K. and Wang, Y. applied numerical theory in proximate analysis and got excellent results. In 1996, Fang, K.T. and Wang, Y. expounded the number theoretic methods applied in statistics [14]. Those works improved the efficiency of algorithms in statistics remarkably. Thus, based on these studies above, we apply number theoretic methods to the MLE of the parameters of K-distribution and use the sequential approach to accelerate the rate of convergence. We also give the way to determine the boundary of initial searching area.

3.2.   The Sequential Number Theoretic Net for Optimization

In the number theoretic methods, there are many kinds of optimal algorithms. The sequential number theoretic net for optimization is an efficient one which is called SNTO for short. We will represent the algorithm in detail.

At first, the point set called glp (good lattice point) set needs to be found in n-dimensional space. The integer vector $\mathbf{h}_k = (h_{k1}, h_{k2}, \cdots, h_{kn}), k = 1, ..., M$ is given, where M is the vector numbers of the glp set, $1 \leq h_{ki} \leq N, h_{ki} \neq h_{kj}, (i \neq j)$ and the greatest common

divisor of M and h_{ki} is one. Thus the set is called the lattice set spanned by vector $\mathbf{h}_k$ which are described as follows,

$$\mathbf{h}_k = (\{\frac{2kh_1 - 1}{2M}\}, \{\frac{2kh_2 - 1}{2M}\}, \cdots, \{\frac{2kh_n - 1}{2M}\}), k = 1, ..., M \tag{4}$$

where {x} denotes the decimal part of x and the lattice point set obtained by (4) is called glp set mentioned above.

On the basis of glp set, we use n-dimensional vector $\mathbf{h}_k$ as the parameters values of the log-likelihood function $L_N(\theta_1, ..., \theta_n; \mathbf{x})$, where θ_i is the i-th parameter. The vector $\mathbf{h}_{max}$ which maximizes $L_N(\mathbf{h}_k; \mathbf{x}), k = 1, ..., M$ is the parameters estimated. For K-distribution, there are only 2 parameters need to be evaluated. So we apply the SNTO algorithm in 2-d space $D^{(t)}$ which is a rectangular area $[\mathbf{a}^{(t)}, \mathbf{b}^{(t)}] = [a_1^{(t)}, b_1^{(t)}] \times [a_2^{(t)}, b_2^{(t)}]$, t is the iterative times, and the log-likelihood function $L_N(a, v; \mathbf{x})$ is expressed as (2).

Step 0 Initializing 2-d space to where we'll map the glp set: let $t = 0, D^{(0)} = [\mathbf{a}^{(0)}, \mathbf{b}^{(0)}]$. We'll represent the way to confirm $D^{(0)}$ in next section;

Step 1 Making number theory net: $P^{(t)}$ is the n_t points glp set which is made by (4) and distributed uniformly to the area $[\mathbf{a}^{(t)}, \mathbf{b}^{(t)}]$ through linear mapping. Usually, let $n_1 > n_2 = n_3 = \cdots$;

Step 2 Searching for the parameters estimated in this step: find $\mathbf{h}_{max}^{(t)} \in P^{(t)} \cup \{\mathbf{h}^{(t-1)}\}$ which maximizes $L_N(a, v; \mathbf{x})$, where $\mathbf{h}^{(-1)} = \phi$ is an empty set;

Step 3 Stopping principle:

given $\delta > 0$, let $\mathbf{c}^{(t)} = \frac{\mathbf{b}^{(t)} - \mathbf{a}^{(t)}}{2}$. If $\max \mathbf{c}^{(t)} < \delta$, it indicates the searching area $D^{(t)}$ is small enough, $\mathbf{h}_{max}^{(t)}$ in this step can be taken as the numerical solution to meet the requirement of accuracy. Then we can stop now. However, if $\max \mathbf{c}^{(t)} \geq \delta$, go to the next step.

Step 4 Shrinking the searching area:

a new area $D^{(t+1)} = [\mathbf{a}^{(t+1)}, \mathbf{b}^{(t+1)}]$ is defined as following

$$\begin{aligned} \mathbf{a}^{(t+1)} &= \max(\mathbf{h}_{max}^{(t)} - \lambda\mathbf{c}^{(t)}, \mathbf{a}^{(t)}) \\ \mathbf{b}^{(t+1)} &= \min(\mathbf{h}_{max}^{(t)} - \lambda\mathbf{c}^{(t)}, \mathbf{b}^{(t)}) \end{aligned} \tag{5}$$

where $i = 1, 2$, λ is the predefined constriction ratio, and usually equal to 0.5. After let $t = t + 1$, we return to step 1.

3.3.   The ways of initializing the searching area in SNTO

An initial rectangular domain in 2-d space is needed at the beginning of SNTO. However, the initial domain may be very large, even be infinite. The moments estimation we calculated should be close to the true values and not hard to be obtained. Therefore we can take the result of MME as the center of the rectangular domain.

The initial rectangular domain is given as following

$$D = [\theta_1^{(0)} - \theta_1^{(0)}, \theta_1^{(0)} + \theta_1^{(0)}] \times [\theta_2^{(0)} - \theta_2^{(0)}, \theta_2^{(0)} + \theta_2^{(0)}]$$
$$= [0, 2\theta_1^{(0)}] \times [0, 2\theta_2^{(0)}]$$

(6)

where $\theta_1^{(0)}, \theta_2^{(0)}$ are the parameters estimated by MME. In this way the left endpoint of the interval can be consistent with the case of $a > 0$ and $v > 0$ as shown in (1).

Another way is to select a relative large rectangular domain and divide it into many small areas. Finally find the MLE by using SNTO on each small area.

4. Simulation Result

Because the method for estimating the parameters of K-distribution using higher order and fractional moments proposed by D. Robert Iskander [15] shows that 0.1 order moment estimator($Y_{0.1}$) is slightly better than that proposed by Raghavan [6] in the case of v close to 0 and using Monte Carlo tests, we have the advantages of knowing the parameters of the generated datasets previously and generating the datasets which are exactly K-distributed, we use Monte Carlo simulation to evaluate and compare the performance with the fractional moments based estimator $Y_{0.1}$.

We will repeat K times the estimation process and get a set of estimates $\hat{v}_i, i = 1, ..., K$ as result.

Moreover, two properties are calculated and compared. One is the normalized bias

$$NBIAS = \frac{|E(\hat{v}_i) - v_0|}{v_0} = \frac{1}{v_0}(|\frac{1}{K}\sum_{i=1}^{K}\hat{v}_i - v_0|)$$

(7)

where v_0 is the true value, and the E(x) is the expectation of x. The other is normalized variance

$$NVAR = \frac{\sigma_{\hat{v}}^2}{v_0^2} = \frac{1}{v_0^2}(\frac{1}{K}\sum_{1}^{K}(\hat{v}_i - v_0)^2) \cdot$$

(8)

A good method should have the normalized bias and the normalized variance both close to 0.

Generally, the shape parameter is often changing from 0.1 which is usual according to the city area to 10 which is according to the plain area in SAR image. The same situation also happens in the sea clutter. So we pay more attention to this parameter range. We divide the range of the shape parameter into three bands, a low band (0.1, 0.2,…, 2.0), a middle band (2, 2.5, …, 10) and a high band (10, 10.5, …, 20). Since in practical application, the number of samples is usually small and limited, we use 256 as the sample size N. Table I gives the parameters used in Monte Carlo trials.

TABLE I.   PAPARAMETERS OF THE MONTE CARLO TRIALS

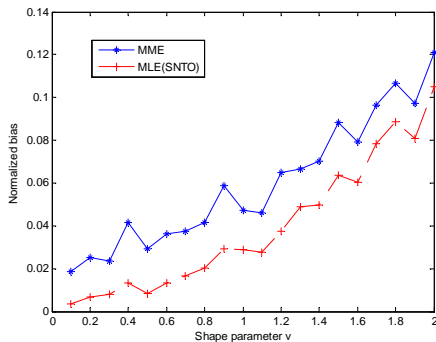| No.	N	K	The Band Of v
1	256	5000	0.1,0.2,---------2
2	256	5000	2,2.5,----------10
3	256	5000	10,10.5,---------20

Fig.1. Normalised bias of the estimated shape parameter in the low band for N=256 and K=5000.
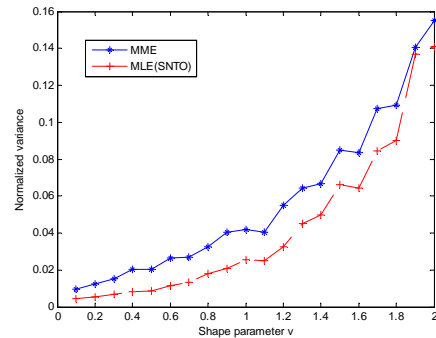

Fig.2. Normalised variance of the estimated shape parameter in the low band for N=256 and K=5000.

Figs. 1 and 2 show the NBIAS and NVAR simulation results in the low band from 5000 Monte Carlo trials. It represents that the MLE (SNTO) achieves much better performance than that of MME. To simplify, we just call MLE (SNTO) as MLE in following. The normalized bias and the normalized variance of MLE are both lower than that of MME.
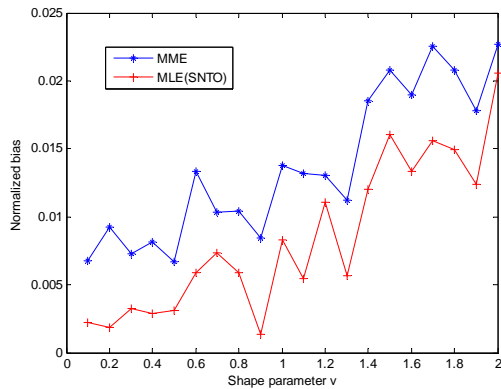

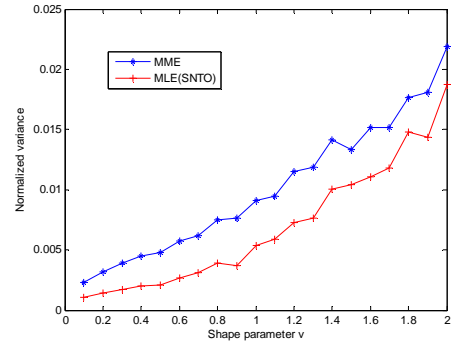Fig.3. Normalised bias of the estimated shape parameter in the low band for N=1000 and K=5000.

Fig.4. Normalised variance of the estimated shape parameter in the low band for N=1000 and K=5000.

We increase the number of samples for both methods with N=1000 and K=5000 and do the simulation shown in Figs. 3 and 4. The increasing of sample size improves the shape parameter estimates and reduces their variance in some extent. Furthermore, both of estimators get more stable results.
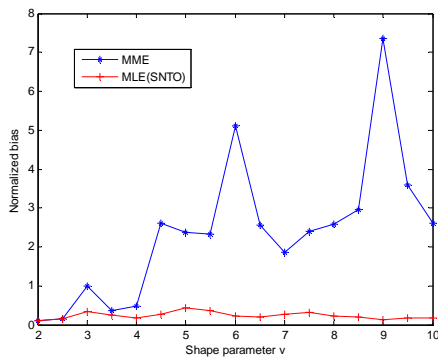

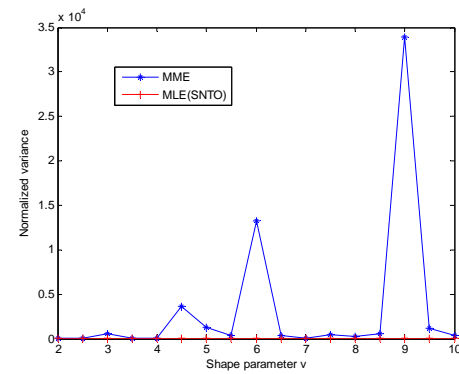Fig.5. Normalised bias of the estimated shape parameter in the middle band for N=256 and K=5000.

Fig.6. Normalised variance of the estimated shape parameter in the middle band for N=256 and K=5000.

In the middle band, the estimation error is becoming larger when the shape parameter getting larger as shown in Figs. 5 and 6. That's because the shape of the K-distribution PDF is less and less sensitive to the value of the shape parameter when the parameter becomes larger. More badly, sometimes the results of MME are illegal and far beyond the reasonable range (0-20). Fortunately, MLE can almost get stable results at any time.
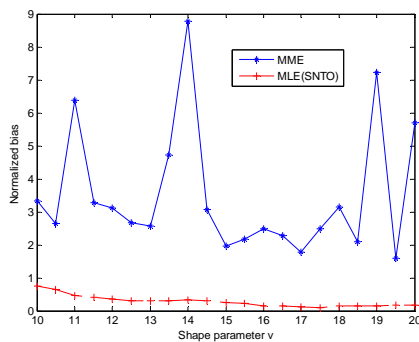

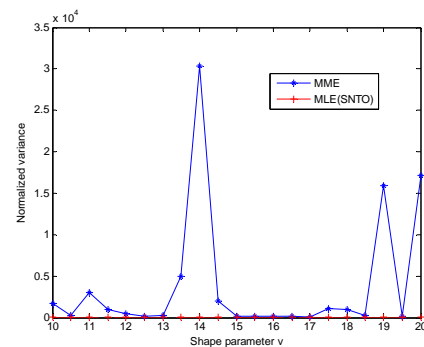Fig.7. Normalised bias of the estimated shape parameter in the high band for N=256 and K=5000.

Fig.8. Normalised bias of the estimated shape parameter in the high band for N=256 and K=5000.

Figs. 7 and 8 show us the high band results of the simulation. It's very similar to the middle band one. There are some illegal values estimated occasionally by MME. The consequence indicates that MLE is more robust than MME.

5. Conclusions

The MLE of the parameters of the K-distribution are difficult to obtain unless computationally expensive numerical methods are used. In this paper, we proposed a new method which uses number theoretic method to calculate the maximum of the log-likelihood function of K-distribution. Its realization is simpler than that of traditional one. For all experiments, we observed that the MLE (SNTO) is robust and superior to MME, even in the situation that the MME gets odd values. But the performance of SNTO is relatively dependent on the initial searching area. With the development of study on this method, more excellent performance could be achieved.

6. References

[1] Jakeman E and Pusey P N. "A model for non-rayleigh sea echo," IEEE Trans. on Antennas Propag, vol. 24, pp. 806-814, 1976.

[2] KATO S, ACKERMAN T P, MATHER J H, "The K-distribution method and correlated-k approximation for a shortwave ra-diative transfer model," Journal of Quantitative Spectros-copy &Radiative Transfer,vol. 62, pp. 109-121, 1999.

[3] Yu Hui ,Wang Yan-fei Yan Hong-hui, "A new method for estimating parameters of K-distributed clutter," Journal of Electronics & Information Technology. pp. 139-142, 2009.

[4] Wang Ying, Mao Er-ke, Han Yue-qi, "Modeling and simulation of correlated K-distributed clutter," SIGNAL PROCESSING,pp. 72-80, 1997.

[5] CONTE E, FARINA A, de MAIO A, "Statistical tests for higher order analysis of radar clutter," IEEE Trans, AES, vol. 41, pp. 205-218, 2005.

[6] Raghavan R S, "A method for estimating parameters of K-Distributed clutter," IEEE Trans on AES, vol. 27, pp. 238-246, 1991

[7] Oliver C J, "Optimum texture estimators for SAR clutter," Journal of Physics : D, vol. 26, pp 1824-1835, 1994.

[8] BlacknellD , Tough R J , "A parameter estimation for the K-Distribution based on zlog(z)," IEE Proc-Radar, Sonar. Navigation, pp. 309-312, 2001.

[9] Wachowiak M P ,RenataSmolíkovÀ, Zurada J M , etal, "Estimation of K Distribution parameters using neural networks," IEEE Trans on Biomedical Engineering, vol. 49, pp. 617-620, 2002.

[10] Joughin I R , Percival D B , Winebrenner D P., "Maximum likelihood estimation of K-distribution parameters for SAR Data," IEEE Trans on Geosciences and Remote Sensing, vol. 31, pp. 989-999, 1993.

[11] Blacknell D. "Comparison of parameter estimators for K-distribution," IEE Proc-Radar, Sonar, Navigation, 1994.

[12] I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K-distribution parameters for SAR data," IEEETrans. Geosciences and Remote Sensing, vol. 32, pp. 989–999, Sept. 1993.

[13] William J. J. Roberts and SadaokiFurui, "Maximum likelihood estimation of K-distribution parameters via the expectation–maximization algorithm," IEEE Trans on Signal Processing, vol. 48, December 2000.

[14] Fang, K.T., Wang, Y. Number-Theoretic Methods in Statistics[M]. Chapman and Hall,1994.

[15] D. Robert Iskander and Abdelhak M. Zoubir, "Estimating the parameters of the K-distribution using ML/MOM approach," IEEE TENCON - Digital Signal Processing Applications, pp. 769–774,1996