Maximum Likelihood Estimation of K-Distribution Parameters via the Expectation–Maximization Algorithm

William J. J. Roberts and Sadaoki Furui, *Fellow, IEEE*

***Abstract*—Maximum likelihood (ML) estimates of K-distribution parameters are derived using the expectation maximization (EM) approach. This approach demonstrates computational advantages compared with 2-D numerical maximization of the likelihood function using a Nelder–Mead approach. For large datasets, the EM approach yields more accurate estimates than those of a non-ML estimation technique.**

***Index Terms*—EM algorithm, K-distribution parameter estimation, maximum likelihood.**

I. INTRODUCTION

THE K-distribution [1] is widely applied to radar signal processing problems and has particular applicability to synthetic aperture radar (SAR) processing [2]. A common requirement is to estimate K-distribution parameters from data samples. Estimates with desirable properties are obtained by applying the ML criterion [3]. Unfortunately, with one exception, analytic equations for ML estimates are not known. The exception was recently demonstrated by Iskander *et al.*, who derive a ML solution for one parameter of the K-distribution using the generalized Bessel function K-distribution [4]. Prior to this, ML estimates have been obtainable only by a two-dimensional (2-D) numerical maximization of the likelihood function [5], [6].

As ML K-distribution parameter estimates are difficult to obtain, research into K-distribution estimation has proposed approximates and alternatives to ML estimation. Ragavan's method [7] equates the ratio of arithmetic and geometric means of K-distributed data to its expected value. The resulting equation can be solved by a numerical search to yield the parameter estimates. In [8], an approximate asymptotic form of the K-distribution is derived, and equations for the ML estimates of this form are obtained. These equations are asymptotically equivalent to the equations for the ML estimates of a gamma distribution. In [9], a modified form of the equations resulting from [8] is proposed based on the results of a neural network analysis. In [10], an approximation based on the gamma distribution is employed, and equations are obtained via matching gamma and K-distribution moments. In [5] and [6], estimation using the method of moments is employed. In all reported simulation studies that we know of (e.g., [5]–[10]), the accuracy of 2-D numerical ML maximization techniques is superior to that of alternative non-ML or approximate ML techniques for large data sizes. However, ML estimates via 2-D numerical maximization require large amounts of computation.

In this paper, we use the expectation maximization (EM) algorithm [11] to derive ML estimates for the K-distribution. In Section II, we derive an iterative solution with the guarantee that the likelihood of estimates produced by successive iteration increases until a stationary point is reached. Each iteration of the EM algorithm requires the solution of equations similar to those for ML gamma distribution parameter estimation, except in this case, the data terms are weighted by functions of the previous parameter estimates and the data. In Section III, we discuss implementation and demonstrate the performance of our solution on K-distributed data.

II. MAIN RESULT

We require the ML estimate $\hat{\lambda}$ of the parameter set λ of the K-distribution from a sequence of independent and identically distributed (*iid*) observations $y = \{y_t, t = 1, \ldots, T\}$, $y_t \in \Re$, i.e.,

$$\hat{\lambda} = \arg\max_{\lambda} p(y|\lambda) = \arg\max_{\lambda} \prod_{t=1}^{T} p(y_t|\lambda) \tag{1}$$

where $p(y_t|\lambda)$ is the probability density function (pdf) of the K-distribution. This pdf may be derived by considering a generalized Rayleigh distribution with a gamma-distributed mean parameter. Other methods of deriving the distribution are possible (see, e.g., [6] and [12]). We may write

$$p(y_t|\lambda) = \int_0^{\infty} p(y_t|w_t)p(w_t|\lambda)\, dw_t \tag{2}$$

and we assume that the sequence $w = \{w_t, t = 1, \cdots, T\}$ is also *iid*. The K-distribution results when

$$p(y_t|w_t) = \frac{2y_t^{N-1}}{(2w_t)^{N/2}\Gamma(N/2)} \exp\left(\frac{-y_t^2}{2w_t}\right) \tag{3}$$

Manuscript received May 6, 1999; revised August 11, 2000. The associate editor coordinating the review of this paper and approving it for publication was Prof. A. M. Zoubir.

W. J. J. Roberts was with the Tokyo Institute of Technology, Tokyo, Japan. He is now with the Information Technology Division, Defence Science Technology Organization, Salisbury, Australia (e-mail: bill.roberts@dsto.defence.gov.au).

S. Furui is with the Computer Science Department, Tokyo Institute of Technology, Tokyo, Japan (e-mail: furui@cs.titech.ac.jp; www.furui.cs.titech.ac.jp).

Publisher Item Identifier S 1053-587X(00)10147-3.

i.e., the generalized Rayleigh distribution resulting from the square root of the sum of the squares of N *iid* Gaussian scalar random variables with zero mean and w_t variance, and

$$p(w_t|\lambda) = \frac{\sigma^{\alpha} w_t^{\alpha-1}}{\Gamma(\alpha)} \exp(-\sigma w_t) \tag{4}$$

i.e., the gamma distribution with parameters $\{\sigma, \alpha\}$, where $\Gamma(\cdot)$ is the gamma function. Substituting (4) and (3) into (2) and performing the integration using [13, p. 340, eq. 3.471(9)] yields

$$p(y_t|\lambda) = \frac{y_t^{\alpha-1+N/2} 2^{2-N/4-\alpha/2} \sigma^{N/4+\alpha/2}}{\Gamma(N/2)\Gamma(\alpha)} K_{\alpha-N/2}(\sqrt{2\sigma} y_t) \tag{5}$$

where $K_\eta(\cdot)$ is given by [13, p. 952, eq. (8.407)] and is known as the modified Bessel function of the second kind of order η. Equation (5) is the K-distribution with parameter set $\lambda = \{\sigma, \alpha\}$. As in other studies, we do not consider here the estimation of N as this parameter corresponds to the number of looks in SAR applications and is generally known.

As there is no known closed-form solution to (1), we consider the EM technique that results in a new estimate $\hat{\lambda}$ of the parameter set, given a current estimate. The properties of the EM algorithm guarantee that until a stationary point is reached, this new estimate has greater likelihood than the current estimate. The equations for the EM are derived by maximization of the auxiliary function [11]

$$\hat{\lambda} = \arg\max_{\lambda} \int_w p(w|y, \lambda') \log p(y, w|\lambda)\, dw \tag{6}$$

where λ' is the current estimate. The convergence of the sequence of these parameter re-estimates is discussed in [14]. In the Appendix, we derive the following equations for the maximizing values of $\hat{\alpha}$ and $\hat{\sigma}$ of (6):

$$\hat{\alpha} \exp(-\Psi(\hat{\alpha})) = \rho \tag{7}$$

$$\frac{1}{\hat{\sigma}} = \frac{1}{T\hat{\alpha}} \sum_{t=1}^{T} a_t y_t \tag{8}$$

where

$$\rho = \frac{\frac{1}{T}\sum_{t=1}^{T} a_t y_t}{\left(\prod_{t=1}^{T} g_t y_t\right)^{1/T}} \tag{9}$$

is the ratio of the arithmetic and geometric means of weighted data, $\Psi(\cdot)$ is the digamma function [13, p. 943] defined as the derivative of the log of the gamma function, and the weights a_t and g_t are given by

$$a_t = \frac{K_{N/2-\alpha'-1}(\sqrt{2\sigma'} y_t)}{\sqrt{2\sigma'} K_{N/2-\alpha'}(\sqrt{2\sigma'} y_t)} \tag{10}$$

$$g_t = \frac{1}{\sqrt{2\sigma'}} \exp\left(-\frac{\partial}{\partial \alpha} \log K_\alpha(\sqrt{2\sigma'} y_t)\Big|_{\alpha=N/2-\alpha'} \right) \tag{11}$$

where σ' and α' are current estimates.

Equations (7)–(11) constitute the iterative EM procedure for the estimation of K-distribution parameters. The iterations may be started with any suitable values of α' and σ'. The iterations may be ceased once convergence criteria are satisfied.

Equations (7) and (8) are of the same form as the equations for the ML estimates of a gamma distribution (see, e.g., [7, Eq. (4a) and (b)]), where now, the data terms are multiplied by a_t or g_t. An explicit solution for $\hat{\alpha}$ is not obtainable directly from (7). We used Newton's method to derive the following iterative solution:

$$\alpha_{n+1} = \alpha_n - \frac{\alpha_n \exp(-\Psi(\alpha_n)) - \rho}{\exp(-\Psi(\alpha_n)) - \alpha_n \Psi'(\alpha_n))} \tag{12}$$

where

α_n	current value of α;
α_{n+1}	new value;
$\Psi'(\cdot)$	trigamma function [15, p. 260].

An appropriate starting point of Newton's method is the the current estimate α'. Once $\hat{\alpha}$ has been obtained, an explicit value of $\hat{\sigma}$ is obtained directly from (8).

III. Simulation Studies

In this section, we detail a simulation of the EM K-distribution estimation scheme derived in Section II. We investigate the performance of the EM scheme and compare it with 2-D numerical ML estimation using the Nelder–Mead (NM) technique [16] and the non-ML method due to Iskander, Zoubir, and Boashash (IZB) [4].

A. *Implementation of the EM Technique*

The implementation was written in Matlab [17] using approximations for the functions $\Psi(\cdot)$, $\Psi'(\cdot)$, and $(\partial/\partial\alpha) \log K_\alpha(\cdot)$, as these functions are not available as part of standard Matlab. The function $\Psi(\cdot)$ was implemented using [15, eqs. (6.3.18) and (6.3.5)]. The $\Psi'(\cdot)$ function was implemented using [15, eqs. (6.4.2) and (6.4.6)]. The $(\partial/\partial\alpha) \log K_\alpha(\cdot)$ function was approximated using

$$\frac{\partial}{\partial \alpha} K_\alpha(x) \approx \frac{K_{\alpha+h}(x) - K_{\alpha-h}(x)}{2h} \tag{13}$$

with $h = 10^{-3}$. The accuracy of the approximation of derivatives in this manner depends on the smoothness of the functions concerned. The iterations of the EM are halted once the difference between successive parameter estimates is less than 0.5% or once the maximum number of iterations (300) is reached.

B. *Experiment and Results*

The execution time of the EM technique and the accuracy of its α estimates were compared to those of the IZB and NM approaches. The latter maximization was implemented by the fmins() function of [17]. The initial parameter values for both the EM and NM methods were those produced by the IZB method.

Table I shows, for three values of T, the true α and σ parameter values used to generate the data together with the mean squared error (MSE) over 1000 trials of the α estimate obtained by the EM scheme α_{EM}, the α estimate via the NM routine

TABLE I

MSEs FOR α_{EM} (α ESTIMATED VIA EM ML), α_{NM} (α ESTIMATED VIA NELDER–MEAD OPTIMIZATION), α_{IZB} (α ESTIMATED USING THE METHOD OF ISKANDER *et al.*). AVERAGES FOR $\mathcal{T}_{EM}/\mathcal{T}_{IZB}$ (RATIO OF EM TO IZB COMPUTATION TIME), AND $\mathcal{T}_{NM}/\mathcal{T}_{IZB}$ (RATIO OF NM TO IZB COMPUTATION TIME)

$N = 2$		$\alpha = 0.5$, $\sigma = 1.0$	$\alpha = 1.0$, $\sigma = 2.0$	$\alpha = 1.5$, $\sigma = 3.0$	$\alpha = 2.0$, $\sigma = 4.0$	$\alpha = 2.5$, $\sigma = 5.0$
	MSE α_{EM}	.0014	.015	.058	.183	.427
	MSE α_{NM}	.0014	.015	.058	.183	.427
$T = 250$	MSE α_{IZB}	.0016	.016	.060	.193	.454
	$\mathcal{T}_{EM}/\mathcal{T}_{IZB}$	1.4	3.1	4.5	8.3	14.4
	$\mathcal{T}_{NM}/\mathcal{T}_{IZB}$	4.3	5.3	8.8	10.6	16.3
	MSE α_{EM}	.0007	.0071	.027	.082	.199
	MSE α_{NM}	.0007	.0071	.027	.082	.199
$T = 500$	MSE α_{IZB}	.0008	.0074	.028	.087	.223
	$\mathcal{T}_{EM}/\mathcal{T}_{IZB}$	2.2	3.9	5.3	10.2	17.5
	$\mathcal{T}_{NM}/\mathcal{T}_{IZB}$	5.1	6.2	9.6	13.9	20.1
	MSE α_{EM}	.0004	.0044	.018	.059	.123
	MSE α_{NM}	.0004	.0044	.018	.059	.123
$T = 750$	MSE α_{IZB}	.0005	.0045	.019	.060	.134
	$\mathcal{T}_{EM}/\mathcal{T}_{IZB}$	3.4	4.3	6.3	13.1	23.2
	$\mathcal{T}_{NM}/\mathcal{T}_{IZB}$	6.1	7.2	10.3	17.2	25.9

α_{NM}, and the α estimate obtained by the IZB method α_{IZB}. A comparison of the computation times is also given in Table I, which shows the averages of the ratio EM to IZB computation time $\mathcal{T}_{EM}/\mathcal{T}_{IZB}$ and the ratio of the NM to IZB computation time $\mathcal{T}_{NM}/\mathcal{T}_{IZB}$. The computation times for the NM and EM schemes did not include the time required to obtain the initial parameter values using the IZB scheme.

C. *Discussion*

For all experiments, we observed that the likelihood at each iteration of the EM was greater than that at the previous iteration. Thus, according to our experimental results, the approximations used in Section III-A did not negate the guarantee of the EM technique to produce a sequence of estimates with increasing likelihood.

The MSE of the α estimates over the 1000 trials of the EM and NM were identical within the convergence tolerances of the algorithms. These two techniques search for *potentially* different points. The NM method attempts to find a maximal point, whereas the EM algorithm will find a stationary point only. The fact that the MSEs for the EM and NM were identical demonstrates that the EM technique consistently located maxima rather than other types of stationary points.

The number of iterations required for convergence of the EM technique was related to the true parameter values. At low α values, the EM approach converged quickly, illustrated by its comparable computational time compared with the noniterative IZB approach. At higher values of α, the EM approach was still quicker than the NM approach but was much slower than the IZB method. For these experiments, the α estimates of the EM always had lower MSE than those of the IZB method. However, in experiments using smaller values of T not reported here, the IZB method produced parameter estimates that were practically indistinguishable from those of the ML methods for significantly reduced computation. Thus, we may conclude that the EM approach is suitable for applications involving large amounts of data when the importance of highly accurate estimates is worth the extra computation required.

APPENDIX

From (6), we require

$$\begin{aligned}\hat{\lambda} &= \arg\max_{\lambda} \int_{w} p(w|y,\lambda')\, \log p(y,w|\lambda)\, dw \\ &= \arg\max_{\lambda} \int_{w} p(w|y,\lambda')\, \log[p(w|\lambda)p(y|w)]\, dw \\ &= \arg\max_{\lambda} \sum_{t=1}^{T} \int_{w} p(w|y,\lambda')\, \log p(w_t|\lambda)\, dw \\ &= \arg\max_{\lambda} \sum_{t=1}^{T} \int_{0}^{\infty} p(w_t|y_t,\lambda') \log p(w_t|\lambda)\, dw_t. \qquad (14)\end{aligned}$$

Substituting (4) in (14), differentiating with respect to σ and α, and setting the result to zero yields

$$\frac{1}{\hat{\sigma}} = \frac{1}{\hat{\alpha}T} \sum_{t=1}^{T} \int_{0}^{\infty} p(w_t|y_t,\lambda') w_t\, dw_t \qquad (15)$$

$$\Psi(\hat{\alpha}) = \log \hat{\sigma} + \frac{1}{T} \sum_{t=1}^{T} \int_{0}^{\infty} p(w_t|y_t,\lambda') \log w_t\, dw_t \qquad (16)$$

where $\Psi(\cdot)$ is the digamma function [13, p. 943]. Using [13, p. 340, eq. 3.471(9)], we have

$$\begin{aligned}\int_{0}^{\infty} p(w_t|y_t,\lambda') w_t\, dw_t &= \int_{0}^{\infty} \frac{p(y_t|w_t)p(w_t|\lambda')}{p(y_t|\lambda')} w_t\, dw_t \\ &= \frac{y_t K_{N/2-\alpha'-1}(\sqrt{2\sigma' y_t})}{\sqrt{2\sigma'} K_{N/2-\alpha'}(\sqrt{2\sigma' y_t})} \qquad (17)\end{aligned}$$

and using the relation $\int \log w f(w)\, dw = (\partial/\partial a) \int w^{a} f(w)\, dw|_{a=0}$, we have

$$\begin{aligned}&\int_{0}^{\infty} p(w_t|y_t,\lambda') \log w_t\, dw_t \\ &\quad = \log \frac{y_t}{\sqrt{2\sigma'}} - \frac{\left.\frac{\partial}{\partial\alpha} K_{\alpha}(\sqrt{2\sigma' y_t})\right|_{\alpha=N/2-\alpha'}}{K_{N/2-\alpha'}(\sqrt{2\sigma' y_t})}. \qquad (18)\end{aligned}$$

Using the definitions for the weights given in (10) and (11) and substituting (17) and (18) into (15) and (16) yields (7) and (8).

ACKNOWLEDGMENT

The authors would like to thank L. B. White, Y. Ephraim, and the anonymous reviewers for their useful comments, as well as D. R. Iskander for supplying the code for his estimation routines. Associate Editor A. M. Zoubir coordinated a timely review of the paper, and his efforts are appreciated by the authors.

REFERENCES

[1] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 806–814, Nov. 1976.

[2] K. D. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.

[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. I.
[4] D. R. Iskander, A. M. Zoubir, and B. Boashash, "A method for estimating the parameters of the K-distribution," *IEEE Trans. Signal Processing*, vol. 47, pp. 1147–1151, Apr. 1999.
[5] I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K-distribution parameters for SAR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 989–999, Sept. 1993.
[6] D. Blacknell, "Comparison of parameter estimators for K-distribution," *Proc. Inst. Elect. Eng., Radar, Sonar, Navigat.*, vol. 141, no. 1, pp. 45–52, Feb. 1994.
[7] R. S. Raghavan, "A method for estimating parameters of K-distributed clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, pp. 238–246, Mar. 1991.
[8] C. J. Oliver, "Optimum texture estimation for SAR clutter," *J. Phys. D: Appl. Phys.*, vol. 26, pp. 1825–1835, 1993.
[9] M. Jahangir, D. Blacknell, and R. G. White, "Accurate approximation to the optimum parameter estimate for K-distributed SAR clutter," *Proc. Inst. Elect. Eng., Radar, Sonar, Navigat.*, vol. 143, no. 6, pp. 383–390, Dec. 1996.
[10] C. J. Oliver and P. Lombardo, "Simultaneous mean and texture edge detection in SAR clutter," *Proc. Inst. Elect. Eng., Radar, Sonar, Navigat.*, vol. 143, no. 6, pp. 391–399, Dec. 1996.
[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *J. R. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.
[12] P. Lombardo and C. J. Oliver, "Estimating the correlation properties of K-distributed SAR clutter," *Proc. Inst. Elect. Eng., Radar, Sonar, Navigat.*, vol. 142, no. 4, pp. 167–178, Aug. 1995.
[13] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, fourth ed. New York: Academic, 1980.
[14] C. F. J. Wu, "On the convergence properites of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
[15] M. Abramowitz and I. E. Stegun, *Handbook of Mathematical Functions*. Washington, DC: U.S. Dept. Commerce, Nat. Bur. Stand., 1972.
[16] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optimiz.*, vol. 9, no. 1, pp. 112–147, 1998.
[17] The Math Works, *MATLAB Users Guide*. Natick, MA: The Math Works, Inc., 1997.

William J. J. Roberts received the Ph.D. degree from George Mason University, Fairfax, VA, in 1997.

From 1998 to 1999, he held a post-doctoral position at the Tokyo Institute of Technology, Tokyo, Japan. Currently, he is a Senior Research Scientist within the Information Technology Division, Defence Science and Technology Organization, Salisbury, Australia. He is interested in aspects of the statistical processing of speech, text, and radar.

Sadaoki Furui (F'93) received the B.S., M.S., and Ph.D. degrees from Tokyo University, Tokyo, Japan in 1968, 1970, and 1978, respectively.

After joining the Nippon Telegraph and Telephone Corporation (NTT) Laboratories, Tokyo, in 1970, he worked on speech analysis, speech recognition, speaker recognition, and speech perception. From 1978 to 1979, he was a Visiting Researcher with AT&T Bell Laboratories, Murray Hill, NJ. He was a Research Fellow and the Director of Furui Research Laboratory, NTT Laboratories and is currently a Professor at the Department of Computer Science, Tokyo Institute of Technology. He has authored or coauthored over 300 published articles and books including *Digital Speech Processing, Synthesis and Recognition*.

Dr. Furui has received various awards from several institutions including the IEEE ASSP Society Senior Award. He is a Fellow of the Acoustical Society of America. In 1993, he served as an IEEE SPS Distinguished Lecturer. He has served on the IEEE Technical Committees on Speech and Multimedia Signal Processing and several international advisory boards in the U.S. and Europe. He is a Board member of the International Speech Communication Association and the Acoustical Society of Japan.