Statistica Sinica **10**(2000), 157-171

APPROXIMATE MAXIMUM LIKELIHOOD METHOD FOR FREQUENCY ESTIMATION

Dawei Huang

Queensland University of Technology

Abstract: A frequency can be estimated by few Discrete Fourier Transform (DFT) coefficients, see Rife and Vincent (1970), Quinn (1994, 1997). This approach is computationally efficient. However, the statistical efficiency of the estimator depends on the location of the frequency. In this paper, we explain this approach from a point of view of an Approximate Maximum Likelihood (AML) method. Then we enhance the efficiency of this method by using more DFT coefficients. Compared to 30% and 61% efficiency in the worst cases in Quinn (1994) and Quinn (1997), respectively, we show that if 13 or 25 DFT coefficients are used, AML will achieve at least 90% or 95% efficiency for all frequency locations.

Key words and phrases: Approximate Maximum Likelihood, discrete Fourier transform, efficiency, fast algorithm, frequency estimation, semi-sufficient statistics.

1. Introduction

A model that has been discussed widely in Time Series Analysis and Signal Processing is the following:

$$x_t = Acos(t\omega + \phi) + u_t, \quad t = 0, \ldots, T-1, \tag{1}$$

where $0 < A$, $0 < \omega < \pi$, and $\{u_t\}$ is a stationary sequence satisfying certain conditions to be specified later. The objective is to estimate the frequency ω from the observations $\{x_t\}$. After obtaining an estimator of the frequency, the amplitude A and the phase ϕ can be estimated by a standard linear least squares method.

Two topics have been of interest in the study of this model: estimation efficiency and computational complexity. Following the idea of Best Asymptotic Normal estimation (BAN), the efficiency hereafter is measured by the ratio of the Cramer - Rao Bound (CRB) for unbiased estimators over the variance of the asymptotic distribution of the estimator. Theoretically, Maximum Likelihood Estimation (MLE) is efficient in this sense and has an especially high convergence rate $O_p(T^{-3/2})$. However, simulation results in Rice and Rosenblatt (1988) show that the MLE obtained by using a standard search program may not be efficient for small and middle sample sizes. This may relate to the computational issue. Since the likelihood function changes from its local maximum to a local minimum within a distance π/T, there is no guarantee that numerical procedures like Newton's method will work even if we use the maximizer of Discrete Fourier Transform (DFT) coefficients as the initial value (see Quinn and Fernandes (1991)).

Many other methods have been proposed to obtain a quick solution, see, for example, Pisarenko (1973), Kay and Marple (1981), Kedem (1986). However, most of them only have convergence rate $O_p(T^{-1/2})$. Several methods that aim at both estimation efficiency and computational complexity have been proposed. Starting from an initial value within a distance less than $O(T^{-1/2})$ from ω, the iterated algorithms in Truong-Van (1990) and Quinn and Fernandes (1991) produce more and more precise estimators. After a few iterations the estimator achieves efficiency. Hannan and Huang (1993) introduced an on-line algorithm that invokes a re-initiation technique to produce an efficient estimator without iteration. The re-initiation technique uses a criterion to decide whether a new time window and a new initial value are needed. However, it also needs an initial value for the true frequency.

A natural initial estimator is the maximizer of DFT coefficients, in the sense of moduli, that can be calculated by the Fast Fourier Transform (FFT) algorithm at the so called Fourier frequencies $2j\pi/T$, $j = 0, \ldots, T-1$. Although FFT cannot be performed recursively, rapid developments in hardware make it possible to obtain the maximizer of DFT coefficients in real time. Also, we shall prove later that as T tends to infinity, the probability that the maximizer is within a distance not greater than π/T from the true frequency tends to one.

Thus, a question is whether we can use the maximal DFT coefficient and its neighbors to estimate ω directly and efficiently? Such an idea may go back to Rife and Vincent (1970). However, as pointed by Quinn (1994), the algorithm in Rife and Vincent (1970) cannot produce an estimator with error order $O_p(T^{-3/2})$ for all ω. Algorithms are presented in Quinn (1994, 1997) to achieve this order. The estimator in Quinn (1994) uses two DFT coefficients: the maximal DFT coefficient and one of its two nearest neighbors. By analyzing the representation of DFT coefficients for model (1), an equation is derived and its solution gives an estimator of the frequency. This method is similar to the moment method in general statistical inference. A Central Limit Theorem is obtained and the error has the order $O_p(T^{-3/2})$ for all ω. However, efficiency changes according to the location of the frequency ω in relation to the nearest Fourier frequency. Roughly speaking, the estimator achieves 99% efficiency for the best case when ω is in the middle of two Fourier frequencies, and 30% efficiency for the worst case when ω is a Fourier frequency. To improve this result, three DFT coefficients are used in Quinn (1997) to achieve 61% efficiency for the worst case.

In this paper, we introduce an Approximate Maximum Likelihood (AML) method to explain and improve the methods in Rife and Vincent (1970) and

Quinn (1994, 1997). Since model (1) is nonlinear in the frequency ω, the likelihood function is very complicated. Taking a Fourier transform, we are still unable to get a sufficient statistic for ω. However, the Fourier transform really simplifies the likelihood function in the sense that only few DFT coefficients near the true frequency are significant. Using these DFT coefficients, we can construct an approximation of the true likelihood function and derive an AML Estimator (AMLE) based on maximizing this approximation. More importantly, we can increase the efficiency of this method by using more DFT coefficients to construct the approximate likelihood function. A formula is derived for calculating the efficiency of an AMLE according to the number of DFT coefficients used. In computational aspect, a closed form of the AMLE is derived when two DFT coefficients are involved. Starting from this estimator and using Fisher's method of scoring for parameters (see, for example, Kendall and Stuart (1967, V2, Chap.18)), we obtain a closed form for an estimator based on more DFT coefficients. This estimator has the same asymptotic distribution as that of the AML estimator while the computational cost is basically the same as that in Rife and Vincent (1970) and Quinn (1994, 1997). For example, no matter where the location of the frequency is, 90% or 95% efficient and tractable estimators can be obtained by using 13 or 25 DFT coefficients respectively. Thus, we solve the problem of estimating frequency efficiently and quickly, without any initial value.

Section 2 derives the AML target function. Section 3 proves the asymptotic properties of the maximizer of the target function. Section 4 explains the calculations. Simulation results are given in Section 5.

2. Target Function for AML

Let

$$S(\lambda) = \begin{bmatrix} 1 & 1 \\ e^{-i\lambda} & e^{i\lambda} \\ \vdots & \vdots \\ e^{-i(T-1)\lambda} & e^{i(T-1)\lambda} \end{bmatrix}, \qquad B = \frac{A}{2}\begin{bmatrix} e^{-i\phi} \\ e^{i\phi} \end{bmatrix}, \quad U = \begin{bmatrix} u_0 \\ \vdots \\ u_{T-1} \end{bmatrix}.$$

Then model (1) can be rewritten as

$$X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{bmatrix} = S(\omega) B + U. \tag{2}$$

Consider the log-likelihood function when the noise $\{u_t\}$ is an i.i.d. $N(0, \sigma^2)$ sequence:

$$L(\lambda, \sigma, A, \phi) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\pi}\|X - S(\lambda) B\|^2.$$

Here $\|V\| = \sqrt{V^*V}$ for a vector V and * denotes the complex transpose. To maximize $L(\lambda, \sigma, A, \phi)$ over all four variables, we only need to minimize $Q(\lambda, A, \phi) = \|X - S(\lambda)B\|^2$ over its three variables. Let $\lambda_k = 2k\pi/T$ be the Fourier frequencies and λ_m be the maximizer of DFT coefficients, p and q be nonnegative integers such that $p \le m \le T-q-1$. We divide a $T \times T$ Fourier matrix into two matrices with dimensions $(q-p+1) \times T$ and $(T-q+p-1) \times T$ respectively:

$$F_1 = \frac{1}{\sqrt{T}} \left[e^{ik\lambda_j} \right]_{j=p,p+1,\ldots,q;\, k=0,\ldots,T-1},$$
$$F_2 = \frac{1}{\sqrt{T}} \left[e^{ik\lambda_j} \right]_{j=0,\ldots,p-1,q+1,\ldots,T-1;\, k=0,\ldots,T-1}.$$

Let I be the identity matrix with an appropriate dimension. Since

$$F_1^* F_1 + F_2^* F_2 = [F_1^*, F_2^*] \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = I,$$

we have

$$Q(\lambda, A_1, A_2) = \|F_1 X - F_1 S(\lambda) B\|^2 + \|F_2 X - F_2 S(\lambda) B\|^2. \tag{3}$$

Let $D(\alpha) = \frac{\sin(T\alpha/2)}{T\sin(\alpha/2)}$ be the Dirichlet function. Then

$$\frac{1}{T} \sum_{t=0}^{T-1} e^{it\alpha} = \frac{e^{iT\alpha} - 1}{T(e^{i\alpha} - 1)} = e^{i(T-1)\alpha/2} D(\alpha). \tag{4}$$

so

$$\frac{1}{\sqrt{T}} F_1 S(\lambda) = \begin{bmatrix} e^{i(T-1)(\lambda_{m-p}-\lambda)/2} D(\lambda_{m-p}-\lambda) & e^{i(T-1)(\lambda_{m-p}+\lambda)/2} D(\lambda_{m-p}+\lambda) \\ \vdots & \vdots \\ e^{i(T-1)(\lambda_m-\lambda)/2} D(\lambda_m-\lambda) & e^{i(T-1)(\lambda_m+\lambda)/2} D(\lambda_m+\lambda) \\ \vdots & \vdots \\ e^{i(T-1)(\lambda_{m+q}-\lambda)/2} D(\lambda_{m+q}-\lambda) & e^{i(T-1)(\lambda_{m+q}+\lambda)/2} D(\lambda_{m+q}+\lambda) \end{bmatrix}.$$

Since the true frequency ω should not be far away from λ_m (see Lemma 1 below), we only need to search the estimator of ω in a neighborhood of λ_m. When $|\lambda - \lambda_m| < \pi/T$, $D(\lambda_k + \lambda)$ is not significant, so the second column in the above matrix can be ignored. Similarly we can ignore the second term in (3) since $F_2 S(\lambda)$ is not significant.

Let

$$Z = \frac{1}{\sqrt{T}} F_1 X, \tag{5}$$
$$H(\lambda) = \left[D(\lambda_{m-p} - \lambda) \cdots D(\lambda_m - \lambda) \cdots D(\lambda_{m+q} - \lambda) \right]^*,$$
$$\Psi = diag \left\{ e^{i(T-1)\lambda_{m-p}/2}, \ldots, e^{i(T-1)\lambda_m/2}, \ldots, e^{i(T-1)\lambda_{m+q}/2} \right\}.$$

Then we have

$$\frac{1}{\sqrt{T}} F_1 S(\lambda) B = A_1 e^{-i(T-1)\lambda/2} \Psi H(\lambda) + O\left(T^{-1}\right).$$

Thus to minimize $Q(\lambda, A_1, A_2)$ in (3), approximately, we consider

$$\min_{\lambda_{m-1}<\lambda<\lambda_{m+1}, C} \left| |Z - Ce^{-i(T-1)\lambda/2} \Psi H(\lambda)| \right|. \tag{6}$$

Changing the double minimum for both λ and C in (6) to two iterated minimums and using the standard least squares method, we can show that (6) is equivalent to maximizing the target function $f(\lambda) = |Z^* \Psi H(\lambda)|^2 / \|H(\lambda)\|^2$. Thus, let

$$N(\lambda) = \frac{1}{\|H(\lambda)\|} H(\lambda), \quad Y = \Psi^* Z, \quad R = \operatorname{Re}\{YY^*\};$$

we estimate ω by the maximizer of

$$f(\lambda) = N(\lambda)^* RN(\lambda), \quad \lambda \in [\lambda_{m-1}, \lambda_{m+1}]. \tag{7}$$

We call this maximizer the Approximate Maximum Likelihood Estimator (AMLE).

3. Asymptotic Behavior of AMLE

First of all, we have

Lemma 1. *Assume that $\{u_t\}$ is a purely non-deterministic ergodic stationary series with zero mean and finite variance. Then*

$$\limsup_{T\to\infty} T |\lambda_m - \omega| \leq \pi, \quad a.s. \tag{8}$$

Proof. Using (4), we have

$$\frac{1}{T}\sum_{t=0}^{T-1} e^{it\lambda} x_t$$
$$= \frac{A}{2} e^{i(T-1)(\lambda-\omega)/2} D(\lambda - \omega) + \frac{A}{2} e^{i(T-1)(\lambda+\omega)/2} D(\lambda + \omega) + \frac{1}{T}\sum_{t=0}^{T-1} e^{it\lambda} u_t.$$

Since $\omega > 0$, $D(\lambda + \omega) \to 0$ as $T \to \infty$. Also, according to the lemma in Hannan (1973), the third term in the above formula tends to zero uniformly for all $\lambda \in [0, \pi]$. Further, there must be an integer k such that $\lambda_k \in [\omega - \frac{\pi}{T}, \omega + \frac{\pi}{T}]$.

So,

$$
\begin{aligned}
&\liminf_{T\to\infty}\left|\frac{1}{T}\sum_{t=0}^{T-1}e^{it\lambda_m}x_t\right| \\
&=\liminf_{T\to\infty}\max_{0\le k<T}\left|\frac{1}{T}\sum_{t=0}^{T-1}e^{it\lambda_k}x_t\right| \\
&\ge\frac{A}{2}\liminf_{T\to\infty}\min\left\{D\left(\lambda-\omega\right);\,\omega-\frac{\pi}{T}\le\lambda\le\omega+\frac{\pi}{T}\right\} \\
&=\frac{A}{2}\lim_{T\to\infty}\frac{1}{T\sin\frac{\pi}{2T}}=\frac{A}{\pi}.
\end{aligned}
$$

If there was a subsequence $\{\lambda_m=\lambda_m(T_j),\ j=1,2,\ldots\}$ such that

$$\liminf_{j\to\infty}T_j\left|\lambda_m\left(T_j\right)-\omega\right|>\left(1+\varepsilon\right)\pi \text{ for } \varepsilon>0,$$

then

$$
\begin{aligned}
&\liminf_{T\to\infty}\left|\frac{1}{T}\sum_{t=0}^{T-1}e^{it\lambda_m}x_t\right| \\
&\le\liminf_{j\to\infty}\left|\frac{1}{T_j}\sum_{t=0}^{T_j-1}e^{it\lambda_m(T_j)}x_t\right| \\
&\le\frac{A}{2}\liminf_{j\to\infty}\max\left\{D\left(\lambda-\omega\right);\left|\lambda-\omega\right|\ge\frac{\left(1+\varepsilon\right)\pi}{T_j}\right\} \\
&=\frac{A}{2}\lim_{T_j\to\infty}\frac{1}{T_j\sin\frac{(1+\varepsilon)\pi}{2T_j}}=\frac{A}{\left(1+\varepsilon\right)\pi}.
\end{aligned}
$$

This contradicts the above formula. So (8) holds.

Next, let $\hat{\omega}$ be the maximizer of $f(\lambda)$ on $[\lambda_{m-p},\,\lambda_{m+q}]$.

Lemma 2. *Under the same condition in Lemma* 1*, we have*

$$T\left(\hat{\omega}-\omega\right)\to 0,\ a.s. \tag{9}$$

Proof. Let $E(\lambda)=T^{-1/2}[1\,e^{-i\lambda}\cdots e^{-i(T-1)\lambda}]^*$. Using the notations in the previous section, it follows from (2) and (5) that

$$Z=\frac{A}{2}e^{-i(T-1)\omega/2-i\phi}\Psi H\left(\omega\right)+F_1\left(\frac{A}{2}e^{i\phi}E\left(\omega\right)+\frac{1}{\sqrt{T}}U\right).$$

Let

$$W=\Psi^*F_1\left(\frac{A}{2}e^{i\phi}E\left(\omega\right)+\frac{1}{\sqrt{T}}U\right).$$

Then

$$Y = \Psi^* Z = \frac{A}{2} e^{-i(T-1)\omega/2 - i\phi} H(\omega) + W.$$

Thus,

$$R = \frac{A^2}{4} H(\omega) H^*(\omega) + M, \tag{10}$$

where

$$M = \mathrm{Re}\left\{ \frac{A}{2} \left[e^{-i(T-1)\omega/2 - i\phi} H(\omega) W^* + e^{i(T-1)\omega/2 + i\phi} W H^*(\omega) \right] + WW^* \right\}.$$

It follows from (7) and (10) that

$$\begin{aligned} f(\hat{\omega}) &= \frac{A^2}{4} \left| N(\hat{\omega})^* H(\omega) \right|^2 + N(\hat{\omega})^* M N(\hat{\omega}) \\ &\geq f(\omega) = \frac{A^2}{4} \| H(\omega) \|^2 + N(\omega)^* M N(\omega). \end{aligned}$$

Then,

$$\begin{aligned} 1 &\geq \left| N(\hat{\omega})^* N(\omega) \right|^2 \\ &\geq 1 - \frac{4}{A^2 \| H(\omega) \|^2} \left[N(\hat{\omega})^* M N(\hat{\omega}) - N(\omega)^* M N(\omega) \right]. \end{aligned} \tag{11}$$

However, let ε^2 be the largest singular value of M and $\varepsilon > 0$; then according to the lemma in Hannan (1973), $\|\frac{1}{\sqrt{T}} U\| \to 0$, *a.s.*, as $T \to \infty$. Since $\omega > 0$, by (4), $\|F_1 E(\omega)\| \to 0$. Thus $\|W\| \to 0$ and then $\varepsilon \to 0, a.s.$, as $T \to \infty$. Also,

$$\begin{aligned} & N(\hat{\omega})^* M N(\hat{\omega}) - N(\omega)^* M N(\omega) \\ &= \left[N(\hat{\omega}) - N(\omega) \right]^* M \left[N(\hat{\omega}) + N(\omega) \right] \\ &\leq \varepsilon \| N(\hat{\omega}) - N(\omega) \| \| N(\hat{\omega}) + N(\omega) \| \leq 4\varepsilon. \end{aligned} \tag{12}$$

Further, for all $\lambda \in [\lambda_{m-p}, \lambda_{m+q}]$ and $p + q > 0$, we have

$$\begin{aligned} \frac{4}{\pi^2} &\leq D\left(\frac{\pi}{T}\right)^2 + D\left(-\frac{\pi}{T}\right)^2 \\ &\leq \sum_{j=m-p}^{m+q} D(\lambda_j - \lambda)^2 = \| H(\lambda) \|^2. \end{aligned} \tag{13}$$

So it follows from (11), (12) and (13) that

$$\lim_{T \to \infty} \left| N(\hat{\omega})^* N(\omega) \right| = 1, \ a.s. \tag{14}$$

Next, let $G(\lambda)=2[\frac{\sin T(\lambda_{m-p}-\lambda)/2}{T(\lambda_{m-p}-\lambda)}\cdots\frac{\sin T(\lambda_{m+q}-\lambda)/2}{T(\lambda_{m+q}-\lambda)}]^{*}$. Since

$$\left|D(\lambda)-\frac{\sin(T\lambda/2)}{T\lambda/2}\right|=\left|\frac{\sin(T\lambda/2)}{T\sin(\lambda/2)}\frac{(\lambda/2)-\sin(\lambda/2)}{(\lambda/2)}\right|\leq\frac{\lambda^{2}}{6},$$

we have

$$\sup_{\lambda_{m-p}\leq\lambda\leq\lambda_{m+q}}\|H(\lambda)-G(\lambda)\|\to 0,\ \text{ as } T\to\infty.$$

Thus, using (13) and (14) we have

$$\lim_{T\to\infty}\frac{|G(\hat{\omega})^{*}G(\omega)|}{\|G(\hat{\omega})\|\|G(\omega)\|}=1,\ a.s. \tag{15}$$

Since for any given ω, there is an integer sequence $\{T_j, j=1,2,\ldots\}$ such that $\sin\frac{T_j\omega}{2}\neq 0$, it follows from (15) that there is also a subsequence of $\{T_j, j=1,2,\ldots\}$ that $\sin\frac{T_j\hat{\omega}}{2}\neq 0$. So, without losing generality, we may assume that both $\sin\frac{T_j\omega}{2}$ and $\sin\frac{T_j\hat{\omega}}{2}$ do not vanish. Let

$$V(\lambda)=\left[\tfrac{2(-1)^{m-p+1}}{2(m-p)\pi-T\lambda},\ldots,\tfrac{2(-1)^{m+q+1}}{2(m+q)\pi-T\lambda}\right]^{*}.$$

Then (15) can be rewritten as $\lim_{T\to\infty}\frac{|V(\hat{\omega})^{*}V(\omega)|}{\|V(\hat{\omega})\|\|V(\omega)\|}=1$, $a.s.$ Also, since $\hat{\omega}\in[\lambda_{m-p},\lambda_{m+q}]$, we conclude

$$\lim_{T\to\infty}\frac{V(\hat{\omega})^{*}V(\omega)}{\|V(\hat{\omega})\|\|V(\omega)\|}=1,\ a.s.$$

So, when $T\to\infty$,

$$\begin{aligned}&\|V(\hat{\omega})-V(\omega)-\left[\frac{\|V(\hat{\omega})\|}{\|V(\omega)\|}-1\right]V(\omega)\|^{2}\\ =&\|V(\hat{\omega})-\frac{\|V(\hat{\omega})\|}{\|V(\omega)\|}V(\omega)\|^{2}\\ =&\|V(\hat{\omega})\|^{2}-2\frac{\|V(\hat{\omega})\|}{\|V(\omega)\|}V(\hat{\omega})^{*}V(\omega)+\|V(\hat{\omega})\|^{2}\to 0,\ a.s.,\end{aligned}$$

$$\frac{2T(-1)^{j+1}(\hat{\omega}-\omega)}{(2j\pi-T\hat{\omega})(2j\pi-T\omega)}-\left[\frac{\|V(\hat{\omega})\|}{\|V(\omega)\|}-1\right]\frac{2(-1)^{j+1}}{2j\pi-T\omega}\to 0,\ a.s.,\ m-p\leq j\leq m+q,$$

$$\frac{T(\hat{\omega}-\omega)}{2j\pi-T\hat{\omega}}-\left[\frac{\|V(\hat{\omega})\|}{\|V(\omega)\|}-1\right]\to 0,\ a.s.,\ m-p\leq j\leq m+q.$$

In particular, taking $j=m-p$ and $j=m+q$, we have

$$\frac{T(\hat{\omega}-\omega)}{2(m-p)\pi-T\hat{\omega}}-\frac{T(\hat{\omega}-\omega)}{2(m+q)\pi-T\hat{\omega}}\to 0,\quad a.s.,\quad T\to\infty.$$

Thus, since $p+q>0,(9)$ holds.

Now we can establish a Central Limit Theorem for the maximizer $\hat{\omega}$. We call a stationary sequence regular if its tail σ-algebra is trivial (see Hannan (1979)). It is known that the regular condition implies the ergodic condition.

Theorem 3. *Suppose that $\{u_t\}$ is a regular stationary sequence with zero mean, its best linear prediction for u_t is the best prediction and its spectral density function $\xi(\lambda)$ is continuous at ω. Let*

$$S_T=\frac{A}{\sqrt{2}T}\|H'(\omega)-\frac{H(\omega)^*H'(\omega)}{\|H(\omega)\|^2}H(\omega)\|. \tag{16}$$

Then

$$T^{3/2}S_T(\hat{\omega}-\omega)\to N(0,2\pi\xi(\omega)).$$

Proof. First we have

$$\hat{\omega}=\omega-\frac{f'(\omega)}{f''(\zeta)},\quad |\zeta-\omega|\le|\hat{\omega}-\omega|. \tag{17}$$

Consider $f'(\omega)$. Since $N(\omega)^*H(\omega)\ge N(\lambda)^*H(\omega)$ for all λ, $N'(\omega)^*H(\omega)=0$. Then, it follows from (10) that

$$f'(\omega)=2N'(\omega)^*\left[A\|H(\omega)\|\mathrm{Re}\left\{e^{i\phi+i(T-1)\omega/2}W\right\}+\mathrm{Re}\{WW^*\}N(\omega)\right].$$

Using the result in Hannan (1979), we can show that $\mathrm{Re}\{e^{i\phi+i(T-1)\omega/2}U\}\to N(0,\pi\xi(\omega)I)$. Then, since $\|F_1E(\omega)\|\to0$, we have

$$\sqrt{T}\mathrm{Re}\left\{e^{i\phi+i(T-1)\omega/2}W\right\}\to N(0,\pi\xi(\omega)I). \tag{18}$$

Also, let

$$K(\omega)\equiv-\|H(\omega)\|\frac{d}{d\omega}\left(\frac{1}{\|H(\omega)\|}\right)=\frac{H'(\omega)^*H(\omega)}{\|H(\omega)\|^2}.$$

Then

$$N'(\omega)=\frac{1}{\|H(\omega)\|}\left[H'(\omega)-K(\omega)H(\omega)\right]. \tag{19}$$

So, we have

$$\frac{\sqrt{2T}f'(\omega)}{A\|H'(\omega)-\frac{H'(\omega)^*H(\omega)}{\|H(\omega)\|^2}H(\omega)\|}\to N(0,2\pi\xi(\omega)). \tag{20}$$

Next, we prove that

$$T^{-2}\left[f''(\zeta)-f''(\omega)\right]=o_p(1). \tag{21}$$

It follows from (7) that $f''(\lambda) = 2N''(\lambda)^* RN(\lambda) + 2N'(\lambda)^* RN'(\lambda)$, $f'''(\lambda) = 2N'''(\lambda)^* RN(\lambda) + 6N''(\lambda)^* RN'(\lambda)$. By (4), $|\frac{d^k}{d\lambda^k} D(\lambda)| = O(T^k)$. Then we have from (13) that

$$\| \frac{d^k}{d\lambda^k} N(\lambda) \| = \| \frac{d^k}{d\lambda^k} \left[\frac{1}{\| H(\lambda) \|} H(\lambda) \right] \| = O\left(T^k\right) \text{ for all } \lambda. \tag{22}$$

Thus, since $f''(\zeta) - f''(\omega) = f'''(\zeta_1)(\zeta - \omega)$, $|\zeta_1 - \omega| \le |\zeta - \omega| \le |\hat{\omega} - \omega|$, it follows from (9) that (21) holds.

Further, manipulating $N'(\omega)$ and $N''(\omega)$ and using (10), we have

$$\frac{f''(\omega)}{T^2} = \frac{A^2}{2T^2} \left[\frac{\left[H'(\omega)^* H(\omega)\right]^2}{\| H(\omega) \|^2} - \| H'(\omega) \|^2 \right] + \frac{2}{T^2} \left[N''(\omega)^* MN(\omega) + N'(\omega)^* MN'(\omega) \right].$$

By (22) and that $M = o_p(1)$, also notice that

$$\| H'(\omega) - \frac{H'(\omega)^* H(\omega)}{\| H(\omega) \|^2} H(\omega) \|^2 = \| H'(\omega) \|^2 - \frac{\left[H'(\omega)^* H(\omega)\right]^2}{\| H(\omega) \|^2},$$

and

$$\frac{f''(\omega)}{T^2} = -\frac{A^2}{2T^2} \| H'(\omega) - \frac{H'(\omega)^* H(\omega)}{\| H(\omega) \|^2} H(\omega) \|^2 + o_p(1). \tag{23}$$

Thus, (16) follows from (17), (20), (21) and (23).

Remark 1. Let $\delta = \frac{T\omega}{2\pi} - m$. When $p = 0$ and $q = 1$, we have

$$H(\omega) = \begin{bmatrix} \frac{\sin(\pi\delta)}{\pi\delta} \\ \frac{\sin(\pi\delta)}{\pi(1-\delta)} \end{bmatrix} + O\left(\frac{1}{T^2}\right),$$

$$T^{-1} H'(\omega) = \begin{bmatrix} \frac{\cos(\pi\delta)}{\pi\delta} - \frac{\sin(\pi\delta)}{(\pi\delta)^2} \\ \frac{\cos(\pi\delta)}{\pi(1-\delta)} + \frac{\sin(\pi\delta)}{\pi^2(1-\delta)^2} \end{bmatrix} + O\left(\frac{1}{T^2}\right).$$

Substituting these into (16), one can verify that

$$\frac{2\pi\xi(\omega)}{S_T^2} = \frac{16\pi^5 d^2 (1 - |\delta|)^2 (2\delta^2 - 2|\delta| + 1)\, \xi(\omega)}{A^2 \sin^2(\pi\delta)} + O\left(\frac{1}{T^2}\right).$$

So, when two DFT coefficients are used, the "asymptotic variance" $\frac{2\pi\xi(\omega)}{T^3 S_T^2}$ in Theorem 1 is asymptotically equal to that in Quinn (1994). As in Quinn (1994), the minimum of this quantity is $\frac{\pi^5 \xi(\omega)}{2A^2}$, while the maximum is $\frac{16\pi^3 \xi(\omega)}{A^2}$. Also, when $p = q = 1$, we have

$$\frac{2\pi\xi(\omega)}{S_T^2} = \frac{8\pi^5 \delta^2 (1 - \delta^2)^2 (3\delta^4 + 1)\, \xi(\omega)}{A^2 (3\delta^4 + 4\delta^2 + 1) \sin^2(\pi\delta)} + O\left(\frac{1}{T^2}\right).$$

Then, when three DFT coefficients are used, the "asymptotic variance" is asymptotically equal to that in Quinn (1997).

Since $\|H'(\omega) - \frac{H'(\omega)^* H(\omega)}{\|H(\omega)\|^2} H(\omega)\|$ is the norm of the regression error of the vector $H'(\omega)$ onto $H(\omega)$, it increases as the dimension $p+q$ increases. So for all $p+q>0$,

$$\frac{1}{T}\|H'(\omega) - \frac{H'(\omega)^* H(\omega)}{\|H(\omega)\|^2} H(\omega)\| \geq \frac{1}{2\pi} > 0. \tag{24}$$

Also, an interesting problem is the relationship between the efficiency of AMLE and the number of DFT coefficients used. We denote the efficiency of AMLE as

$$\eta(\omega) \equiv \frac{\frac{48\pi\xi(\omega)}{T^3A^2}}{\frac{2\pi\xi(\omega)}{T^3S_T^2}} = \frac{12\|H'(\omega) - \frac{H'(\omega)^* H(\omega)}{\|H(\omega)\|^2} H(\omega)\|^2}{T^2}. \tag{25}$$

Figure 1 shows $\eta(\omega)$ for 2, 3, 13 and 25 DFT coefficients. The X-axis is the distance between the true frequency and λ_m in the unit π/T. It is clear that the minimum of $\eta(\omega)$ occurs at λ_m. Also, when $\omega=\lambda_m$ and $p=q$, $H'(\omega) = \left[-\frac{\cos p\pi}{2\sin p\pi/T} \cdots 0 \cdots \frac{\cos p\pi}{2\sin p\pi/T}\right]$, $H(\omega)^* H'(\omega) = 0$. Thus

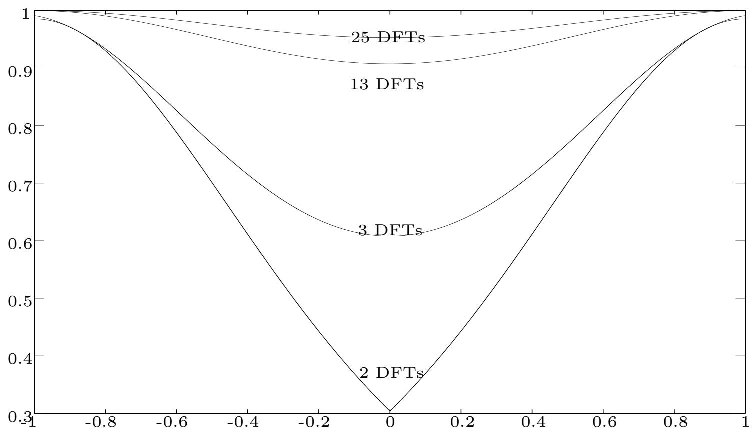

Figure 1. Efficiency curve of AML, 2, 3, 13 and 25 DFT coefficients are used.

Corollary 4. *Under the conditions in Theorem* 3 *and when* $p=q$,

$$\min_{0<\omega<\pi} \eta(\omega) \geq \frac{6}{T^2}\sum_{j=1}^{p} \frac{1}{\sin^2 \frac{j\pi}{T}} > \frac{6}{\pi^2}\sum_{j=1}^{p}\frac{1}{j^2}.$$

Notice that $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$ (see, for example, Dym and McKean (1972, Chap. 2)), the efficiency approaches one very quickly. For example, we can achieve 95% efficiency, no matter where ω is, when $p=12$.

4. Algorithm

In this section, we use a subscript to indicate the number of DFT coefficients on which the statistics are based. Subscript 0 means two DFT coefficients while subscript $p > 0$ means (2p+1) DFT coefficients corresponding to the Fourier frequencies $\lambda_j, j = m - p, \ldots, m + p$. For example,

$$H_p(\lambda) = \left[D(\lambda_{m-p} - \lambda) \cdots D(\lambda_{m+p} - \lambda) \right]^*, \quad f_p(\lambda) = N_p(\lambda)^* R_p N_p(\lambda).$$

First consider estimation based on two DFT coefficients. One of them is the maximal DFT coefficient corresponding to λ_m. We choose the other by the sign of $f_1'(\lambda_m)$: let

$$j = \begin{cases} 1, & f_1'(\lambda_m) > 0; \\ -1, & f_1'(\lambda_m) \leq 0. \end{cases}$$

Then the target function is $f_0(\lambda) = N_0(\lambda)^* R_0 N_0(\lambda)$ with $R_0 = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}$,

$$r_{11} = \left| \sum_{t=0}^{T-1} e^{it\lambda_m} x_t \right|^2, \quad r_{22} = \left| \sum_{t=0}^{T-1} e^{it\lambda_{m+j}} x_t \right|^2,$$
$$r_{12} = \mathrm{Re}\left[e^{-i\frac{T-1}{2}(\lambda_{m+j} - \lambda_m)} \sum_{t=0}^{T-1} e^{-it\lambda_m} x_t \sum_{t=0}^{T-1} e^{it\lambda_{m+j}} x_t \right].$$

Let $[\cos\theta, \sin\theta]$ be the eigenvector corresponding to the maximum eigenvalue of R_0. Then

$$\tan\theta = -\kappa = -\frac{r_{11} - r_{22} - \sqrt{(r_{11} - r_{22})^2 + 4r_{12}^2}}{2r_{12}}.$$

So, we derive a closed form for $\hat{\omega}_0$:

$$\frac{D(\lambda_{m+j} - \hat{\omega}_0)}{D(\lambda_m - \hat{\omega}_0)} = \frac{\sin\frac{\lambda_m}{2}\cos\frac{\hat{\omega}_0}{2} - \cos\frac{\lambda_m}{2}\sin\frac{\hat{\omega}_0}{2}}{\sin\frac{\lambda_{m+j}}{2}\cos\frac{\hat{\omega}_0}{2} - \cos\frac{\lambda_{m+j}}{2}\sin\frac{\hat{\omega}_0}{2}} = -\kappa,$$
$$\hat{\omega}_0 = 2\arctan\frac{\sin\frac{\lambda_m}{2} - \kappa\sin\frac{\lambda_{m+j}}{2}}{\cos\frac{\lambda_m}{2} - \kappa\cos\frac{\lambda_{m+j}}{2}}. \tag{26}$$

Now, for any integer $p > 0$, let

$$\bar{\omega}_p = \hat{\omega}_0 - \frac{f_p'(\hat{\omega}_0)}{f_p''(\hat{\omega}_0)}. \tag{27}$$

Although $\bar{\omega}_p$ may not be exactly the same as the maximizer $\hat{\omega}_p$ of $f_p(\lambda)$, we have

Theorem 5. *Under the same conditions as in Theorem* 3*,*

$$T^{1.5}S_p\left(\bar{\omega}_p-\omega\right)\rightarrow N\left(0,2\pi\xi\left(\omega\right)\right), \tag{28}$$

where $S_p=\frac{A}{\sqrt{2T}}\|H_p'\left(\omega\right)-\frac{H_p'(\omega)^*H_p(\omega)}{|H_p(\omega)|^2}H_p\left(\omega\right)\|$.

Proof. Since

$$\begin{aligned}\bar{\omega}_p-\omega&=\hat{\omega}_0-\omega-\frac{f_p'\left(\hat{\omega}_0\right)-f_p'\left(\omega\right)+f_p'\left(\omega\right)}{f_p''\left(\hat{\omega}_0\right)}\\&=\left[1-\frac{f_p''\left(\zeta\right)}{f_p''\left(\hat{\omega}_0\right)}\right]\left(\hat{\omega}_0-\omega\right)-\frac{f_p'\left(\omega\right)}{f_p''\left(\hat{\omega}_0\right)},\quad\left|\zeta-\omega\right|\leq\left|\hat{\omega}_0-\omega\right|,\end{aligned}$$

we have

$$T^{1.5}S_p\left(\hat{\omega}_0-\omega\right)=O_p\left(1\right),\quad 1-\frac{f_p''\left(\zeta\right)}{f_p''\left(\hat{\omega}_0\right)}=o_p\left(1\right).$$

Now (28) follows by the argument in the proof of Theorem 3.

Remark 2. In practice, when the sample size is small and the Signal to Noise Ratio (SNR) is low, it is better to modify (27) by adding one more step:

$$\bar{\omega}_1=\hat{\omega}_0-\frac{f_p'\left(\hat{\omega}_0\right)}{f_p''\left(\hat{\omega}_0\right)},\quad\bar{\omega}_p=\bar{\omega}_1-\frac{f_p'\left(\bar{\omega}_1\right)}{f_p''\left(\bar{\omega}_1\right)}.$$

Remark 3. To reduce the computational complexity while keeping the same asymptotic property, we can replace $H_p\left(\lambda\right)$ by

$$V_p\left(\lambda\right)=\left[\frac{2(-1)^{p-1}}{2(m-p)\pi-T\lambda}\cdots\frac{2(-1)^{p+1}}{2(m+p)\pi-T\lambda}\right]^*.$$

When two DFT coefficients are used, this leads to

$$\bar{\omega}_0=\frac{\lambda_m-\kappa\lambda_{m+j}}{1-\kappa},$$

which is similar to the estimator in Quinn (1994).

Remark 4. Compared with the existing methods, the computational complexity of this method is acceptable. As in Rife and Vincent (1970) and Quinn (1994, 1997), the major computation is due to FFT and is $\mathrm{O}(T\log T)$. For calculating $\bar{\omega}_p$ based on $V_p\left(\lambda\right)$, one only needs $\mathrm{O}(p)$ extra multiplications and additions. So the calculation of AML is basically the same as that in Rife and Vincent (1970) and Quinn (1994, 1997). Suppose we set out to find the maximizer of the likelihood function (Rice and Rosenblatt (1988)) or the periodogram (Hannan (1973)) by iteration procedures like Newton's method. Even if we start from the maximizer of the periodogram and assume only one iteration is used (though this is far from enough according to the investigations in Quinn and Fernandes (1991) and Quinn (1994)), the computational complexity is proportional to $O(T\log T)$ plus $T\times C$ calculations for the first and second derivatives of log-likelihood functions, or periodograms. The constant C in the second part is a very large number since the derivatives consist of T sine and cosine functions and no fast algorithms are available for the calculation.

5. Simulation

Simulation was done and some results are displayed in Tables 1 to 3. Two frequencies, $\frac{50\pi}{T}$ and $\frac{51\pi}{T}$ with $T=128$, are estimated, SNR($=10\log_{10}\frac{A^2}{\sigma^2}$) changed from -5 to 5, and 2, 3, 13 and 25 DFT coefficients were used ($p=0,1,6,12$). The entries in the tables are the average sample efficiencies of 1000 replications for each case. These sample efficiencies are calculated by the ratio of CRB over the sample mean squares errors. The theoretical efficiency η calculated by (25) is listed on the last column.

Table 1. $T=128$, $\omega=\frac{50\pi}{T}$, 128 DFT coefficients.

p	-5	-4	-3	-2	-1	0	1	2	3	4	5	η
0	.4349	.4286	.4218	.4148	.4078	.4011	.3947	.3886	.3831	.3780	.3733	.3040
1	.5042	.5115	.5184	.5252	.5320	.5387	.5452	.5515	.5576	.5633	.5687	.6080
6	.8962	.9074	.9144	.9185	.9208	.9218	.9221	.9219	.9213	.9206	.9197	.9074
12	.9145	.9275	.9359	.9414	.9450	.9437	.9488	.9497	.9501	.9503	.9502	.9529

Table 2. $T=128$, $\omega=\frac{51\pi}{T}$, 128 DFT coefficients.

p	-5	-4	-3	-2	-1	0	1	2	3	4	5	η
0	.0001	.0029	.0577	.1111	.6376	.9409	.9456	.9486	.9501	.9503	.9490	.9855
1	.0001	.0026	.0564	.1006	.3700	.9420	.9452	.9470	.9477	.9472	.9453	.9912
6	.0001	.0004	.0586	.1112	.9442	.9633	.9626	.9611	.9586	.9549	.9501	.9999
12	.0001	.0025	.0594	.1105	.6708	.9619	.9609	.9591	.9564	.9527	.9477	.9999

In Tables 1 and 2, the standard DFT with 128 coefficients was used. For $\omega=\frac{50\pi}{T}$, which corresponds to the worst case, the results were consistent with the theoretical efficiency. However, for the best case $\omega=\frac{51\pi}{T}$, when SNR <0, a big difference between the efficiencies given by simulation and theory occurred. We found that it is due to the location of λ_m. In this case $|\lambda_m-\omega|$ could be more than π/T. To solve the problem, we calculated DFT on 256 frequencies $\frac{j\pi}{T}, j=0,\ldots,255$. Suppose that $\frac{n\pi}{T}$ is the maximizer. Then let $\lambda_m=\frac{n\pi}{T}$ if n is even; $\lambda_m=\frac{(n+1)\pi}{T}$ if n is odd and $|\sum_{t=0}^{T-1}e^{it(n+1)\pi/T}x_t|>|\sum_{t=0}^{T-1}e^{it(n-1)\pi/T}x_t|$; otherwise $\lambda_m=\frac{(n-1)\pi}{T}$. Based on such a λ_m, we obtained results which were close to the theoretical values in Table 3.

Table 3. $T = 128$, $\omega = \frac{51\pi}{T}$, 256 DFT coefficients.

p	-5	-4	-3	-2	-1	0	1	2	3	4	5	η
0	.0455	.8978	.9132	.9253	.9343	.9409	.9456	.9486	.9501	.9503	.9490	.9855
1	.0456	.9130	.9235	.9315	.9375	.9420	.9452	.9470	.9477	.9472	.9453	.9912
6	.0457	.9625	.9643	.9648	.9647	.9639	.9626	.9606	.9577	.9539	.9490	.9999
12	.0457	.9650	.9647	.9645	.9638	.9626	.9609	.9586	.9566	.9517	.9466	.9999

References

Dym, H. and McKean, H. P. (1972). *Fourier Series and Integrals.* Academic Press, New York.

Hannan, E. J. (1973). The estimation of frequency. *J. Appl. Probab.* **10**, 510-519.

Hannan, E. J. (1979). The central limit theorems for time series regression. *Stoch. Proc. Appl.* **9**, 281-289.

Hannan, E. J. and Huang, D. (1993). On line frequency estimation. *J. Time Ser. Anal.* **14**, 147-161.

Kay, S. M. and Marple, S. L. (1981). Spectrum analysis - a modern perspective. *Proc. IEEE* **69**, 1380-1419.

Kedem, B. (1986). On frequency detection by zero-crossing. *Signal Processing* **10**, 303-306.

Kendall, M. G. and Stuart, A. (1967). *The Advanced Theory of Statistics.* Charles Griffin & Co., London.

Pisarenko, V. F. (1973). The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astr. Soc.* **33**, 347-366.

Quinn, B. R. and Fernandes, J. M. (1991). A fast efficient technique for the estimation of frequency. *Biometrika* **78**, 489-498.

Quinn, B. G. (1994). Estimating frequency by interpolation using Fourier coefficients. *IEEE Trans. Signal Processing* **42**, 1264-1268.

Quinn, B. G. (1997). Estimation frequency, amplitude and phase from the DFT of a time series. *IEEE Trans. Signal Processing* **45**, 814-817.

Rice, J. A. and Rosenblatt, M. (1988). On frequency estimation. *Biometrika* **75**, 477-484.

Rife, D. C. and Vincent, G. A. (1970). Use of the discrete Fourier transform in the measurement of frequencies and levels of Tones. *Bell Syst. Tech. J.* **49**, 197-228.

Truong-Van (1990). A new approach to frequency analysis with amplified harmonics. *J. Roy. Statist. Soc. Ser. B* **52**, 203-221.

School of Mathematics, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia.

E-mail: d.huang@fsc.qut.edu.au

(Received May 1997; accepted May 1999)