Title: Threshold behaviour of Maximum Likelihood Frequency Estimation

Topic number: EAB7

Student Name: Licheng Kong **Student ID: z3372129**

A. Problem statement

As is well-known, with MLE, threshold effect occurs when SNR drops below a specific point in that estimation MSE grow rapidly and the estimates quickly become unreliable. This is mainly due to the occurrence of outliers. It is observed that in the overall outlier probability, single outlier probability is dominant when SNR is in the threshold region. In order to probe deeply into threshold behaviour of MLE, we need to formulate the expression of single outlier probability and investigate its properties. Moreover, new threshold expression is needed for the better accuracy in indicating threshold onset.

B. Objective

Basic objective: to study and compare the MLE threshold expressions in the literature.
Advanced objective 1: to analyse threshold behaviour from the angle of outliers.
Advanced objective 2: to formulate a new threshold expression using knowledge of single outlier or other methods.

C. My solution

Universal expression has been derived for the probability of exactly L number of outliers for the sake of investigating single outlier probability. This is accomplished by using the Rayleigh and Rician distributions of different frequency bins.
A new threshold expression has been formulated using the rational-polynomial method.

D. Contributions (at most one per line, most important first)

New threshold expression for unbiased MLE threshold.
Universal expression for the probability of exactly L number of outliers
The expression and properties of single outlier probability

E. Suggestions for future work

Use Saddle Point Theory or Laplace Asymptotic Method to conquer Qn and Q1
Improve Q&A threshold by relaxing assumptions
Improve Knockaert's threshold by using the tighter lower bounds.

While I may have benefited from discussion with other people, I certify that this report is entirely my own work, except where appropriately documented acknowledgements are included.

Signature: 孔令 [signature] Date: _29_ / _10_ / __2015__

Pointers

List relevant page numbers in the column on the left. Be precise and selective: Don't list all pages of your report!

6	Problem Statement
6	Objective

Theory (up to 5 most relevant ideas)

7	Maximum Likelihood Estimation
8	Discrete Fourier Transformation
9	CRB
10	Threshold effect

Method of solution (up to 5 most relevant points)

25	Statistically modelling the outlier probability concerning Rayleigh and Rice distribution.
34-35	Rational-polynomial approximation

Contributions (most important first)

34-37	New threshold expression for unbiased MLE threshold.
25-26	Universal expression for the probability of exactly L number of outliers
27-32	The expression and properties of single outlier probability

My work

15	Studying and comparing expressions in the literature
25-27	Derivation of outlier probabilities
34-35	Derivation of new threshold expression

Results

19	Comparison of and comments on existing expressions
25-27	Universal and single outlier probability expression
37	Verification and viability analysis of new threshold

Conclusion

39	Statement of whether the outcomes met the objectives
40	Suggestions for future research

Literature: (up to 5 most important references)

41	D.C. Rife and R.R. Boorstyn, 1974
41	E. Aboutanios, 2005
41	L. Knockaert, 1997
41	B.G. Quinn and P.J. Kootsookos, 1994
41	R.C. Williamson et al, 1995

THE UNIVERSITY OF
NEW SOUTH WALES

SCHOOL OF ELECTRICAL ENGINEERING
AND TELECOMMUNICATION

Threshold Behavior of Maximum Likelihood Frequency Estimators

by

Licheng Kong

Thesis submitted as a requirement for the degree
Bachelor of Engineering (Electrical Engineering)

Submitted: October 29, 2015

Student ID: z3372129

Supervisor: Elias Aboutanios

Topic ID: EAB7

Abstract

Frequency estimation with the single tone signal model in Additive White Gaussian Noise (AWGN) is well-known to be plagued with threshold effect in which the growth of noise would severely undermine estimation accuracy. Indicators of the onset of this phenomenon, the thresholds, are thus indispensable in judging whether or not to reject an estimate with respect to the signal to noise ratio and number of samples. In this thesis, we probe into the mechanism behind threshold behaviour from the angle of outlier probabilities. The probability of single outlier is theoretically predicted to be dominant in the overall outlier probability in the threshold region, and is backed by experimental observations. Quinn and Aboutanios' threshold is chosen as the benchmark throughout this thesis because it is shown to be highly accurate. New threshold expression is proposed via rational-polynomial method with both single and total outlier probability. Thresholds provided by the new expression are closer to the theoretical 'knees' of mean square estimation error when the number of samples is within the normal usage range.

Contents

List of Figures

List of Tables

1 Introduction

Frequency and phase estimation of a complex exponential in Additive White Gaussian Noise (AWGN) is a well-studied and fundamental area in digital signal processing and communications. It leads an inspiring exploration into the more complex and practical problems such as data recovery of radio systems [1], the bearing system in radars and sonars [2], the estimation of heart rate in biomedicine [3] and the electricity managing system in the power grids [11]. The maximum likelihood (ML) estimators, which is usually distinguished from the other general estimation method named maximum a posteriori (MAP) estimation, is embracing more popularity in both fields of research and industrial use. The reason lies in its cheaper availability brought by the Fast Fourier Transform (FFT) algorithm and the faster newly available digital processors. This can be exemplified by the substantial reduction in computational complexity when the classical Discrete Fourier Transform (DFT) algorithm upgrades to FFT. The required number of complex multiplications releases from N^2 to only $\frac{N}{2}\log_2 N$ when N entries of data is to be sampled and transformed [12].

Precision requirement on our estimators varies in different circumstances but estimations are considered meaningful only when their mean error is within a tolerable range. In other words, we need to determine whether an estimate is reliable or not by a threshold in terms of signal to noise ratio (SNR) and reject the result when noise is too high. Bad estimations may lead to major malfunctions in a digital processing system. For instance, the pass-band signals in a communication system may not be correctly demodulated to the base-band if the carrier frequency is estimated poorly [1]. For such cases, we demand a benchmark to reasonably reject bad estimations. This requires the knowledge of how estimation error is related to noise level.

As a matter of fact, the estimation error does not vary linearly with the signal to noise ratio and it actually deviates from the linear trend and grow rapidly when SNR drops below a specific point corresponding to the number of samples N. This is the general behavior of threshold effect and choosing a qualified indicator of threshold would have to be based on good knowledge of the behavior of estimation error.

In [5], the Mean Square Error (MSE) of unbiased FFT-based estimation is composed of two types of variances that are weighted by the outlier probability (Q). One is the uniformly distributed noise variance and the other is the Cramer-Rao Lower Bound (CRB). Outliers are defined to occur when the Fourier coefficients at noise frequencies are greater than that at true frequency after FFT. Among these elements involved in estimation error, the key to the mystery of threshold effect lies in the outlier probability since both two types of variances are easy to determine but Q can be more complex to deal with. It is observed that in the probability of all possible numbers of outliers (Q_N), single outlier probability (Q_1) is dominant. This indicates that Q_1 may be able to represent Q_N and also motivates me to investigate single outlier probability and formulate it with statistical models.

After the research on outlier probabilities, we can interpret the threshold with an abundance of knowledge. A Number of threshold expressions have been formulated in the literature. Among them, Quinn [4] and Aboutanios' [15] expression is shown to be very accurate and is chosen as the benchmark throughout this thesis. New expression is also proposed using rational-polynomial approximation. This expression is shown to be closer to the theoretical 3dB deviation points of MSE from CRB when N is within 1024 but suffers from asymptotic deviation as N grows beyond 5820. In fact, usual implementation of FFT-based estimation involves a small number of samples (less than 1024) in avoidance of large latency such that this expression can provide more accurate reference for the reliability estimates.

The rest of this thesis is organized in the following scheme. A brief statement of problem is given in Section 2. Research objectives are specified in Section 3 and Section 4 will provide some background knowledge including MLE, DFT, Maximum Bin Search, CRB and etc. Section 5 will serve to explain the composition of estimation MSE and the role of outlier probability. Three threshold expressions are introduced and compared in Section 6. My research outcomes are included in Section 7 and 8. The former section comprises the statistical modelling of single outlier probability and its capability in representing total outlier probability. The later section presents the newly prosed threshold expression. We conclude by restating the research findings, comparing the results and projecting directions for future study.

2 Problem Statement

Referring again to the phenomenon of threshold effect, with unbiased Maximum Likelihood frequency estimation, when signal to noise ratio drops below a specific point, the MSE grow rapidly and the estimates quickly become unreliable. This is mainly due to the occurrence of outliers. It is observed that in the overall outlier probability, single outlier probability is dominant when SNR is in the threshold region. In order to probe deeply into threshold behaviour of MLE, we need to formulate the expression of single outlier probability and investigate its properties. In addition, new threshold expression is needed for the better accuracy in indicating threshold onset.

3 Research Objectives

The objectives of this thesis consists of one basic objective and two advanced objectives.

• Basic objective: to study and compare the MLE threshold expressions in the literature.

• Advanced objective: to analyse threshold behaviour from the angle of outliers.

• Advanced objective: to formulate a new threshold expression using knowledge of single outlier or other methods.

4 Background Theory

4.1 Signal Model

The signal model below is used for this thesis.

$$x[k] = s[k] + w[k], \quad k = 0, 1, 2, ..., N-1, \tag{1}$$

$$where\ s[k] = Ae^{2\pi k \frac{f}{fs} + \phi}, \quad k = 0, 1, 2, ..., N-1$$

$$and\ w[k] \sim N(0, \sigma^2),$$

where A, f, fs, ϕ, k stand for the signal amplitude, the true frequency, the sampling frequency, the phase shift and the sample index respectively. And w is the Additive White Gaussian Noise (AWGN) which has a zero mean and a variance of σ^2 for both real and imaginary part.

4.2 Estimation Error and SNR

The estimation error is defined as the difference between the true frequency and the estimated frequency. In order to examine the features of an estimator, we usually repeat an estimation for hundreds of thousands of times. The errors are squared and then averaged to produce a more accurate test result, which is termed Mean Square Error(MSE) of ML estimators. One can also use Root Mean Square Error (RMSE) for the same purpose.

$$MSE = \frac{1}{J}\sum_{i=1}^{J}(\hat{f}_i - f)^2, \tag{2}$$

where J is the number of repeated experiments and $\hat{f}$ is the estimated frequency.

In fact, the MSE of an estimator is closely related to the signal to noise ratio. In the signal model given by equation 1, both real and imaginary parts of the AWGN has a variance of σ^2, so signal to noise ratio is defined as

$$\rho = \frac{signalpower}{noisepower} = \frac{A^2}{\sigma^2}. \tag{3}$$

Note that in this thesis, we refer to SNR as signal to noise ratio in decibel, i.e.

$$SNR(dB) = 10log_{10}(\rho). \tag{4}$$

4.3 Maximum Likelihood Estimation

Generally for a defined statistical model, Maximum Likelihood Estimation (MLE) searches for sets of parameters that could maximize the likelihood function of an observation [13].Let $\hat{x} \in X$ be an observation of system X and θ be the set of related parameters such that the likelihood function is expressed as L$(x|\theta)$.Then ML selects $\hat{\theta_{ML}}$ as the parameter set that satisfies

$$L(x|\hat{\theta}_{ML}) = \sup_{\theta} L(x|\theta). \tag{5}$$

In [10], ML is equivalently defined in terms of partial derivative of the log likelihood function $\log L(x|\theta)$ such that $\hat{\theta}_{ML}$ satisfies the equation

$$\left.\frac{\mathrm{d}log(L(x|\theta))}{\mathrm{d}\theta}\right|_{\theta=\hat{\theta}_{ML}} = 0. \tag{6}$$

Therefore, the process of determining $\hat{\theta}_{ML}$ is an ML estimator.

In the context of frequency estimation, it is impractical to search for a frequency that maximizes the DFT coefficient along the spectrum as this would require an infinitely large number of samples [6]. As mentioned earlier, DFT operations are usually implemented using the FFT algorithm for less computational complexity and faster speed. The concept of MLE is usually implemented using FFT-based estimation which consists of two stages, the Maximum Bin Search (MBS) and the fine search stage. The first process of MBS is FFT but only for a limited number of samples and, after mapping the signal to the frequency domain, search for the greatest Fourier coefficient (Note that we also refer to Fourier coefficients as frequency bins). The result of MBS is given by

$$\hat{f} = arg\ max|A(f)|,$$

where A stands for the DFT operator.

After this, different fine search algorithms, such as the Dichotomous Search [14] and the Iterative Search by Interpolation [6], will improve the resolution and achieve very accurate estimates. Since MLE is the mechanism behind FFT-based estimation, we also refer to FFT-based estimation vaguely as MLE as in the topic.

4.4 Discrete Fourier Transform

In this subsection we will briefly look at the operation of DFT followed by the Maximum Bin Search (MBS). We include the concept of zero-padding in MBS stage to have a clearer view of the resolution.

The DFT operator is given by the following equation where $s[k]$ is the sampled pure sinusoid in our signal model.

$$S[n] = \frac{1}{N}\sum_{k=0}^{N-1} s[k]e^{\frac{-j2\pi kn}{N}}, \quad n = 0, 1, 2, ..., N-1. \tag{7}$$

Substituting s[k] into equation (7) and we have

$$\begin{aligned} S[n] &= \frac{1}{N}\sum_{k=0}^{N-1} Ae^{j2\pi k\frac{f}{fs}}e^{\frac{-j2\pi kn}{N}} \\ &= \frac{Ae^{j\phi}}{N}\sum_{k=0}^{N-1} e^{j2\pi(\frac{f}{fs}-\frac{n}{N})^k}, \quad n = 0, 1, 2, ..., N-1. \end{aligned} \tag{8}$$

According to the linearity of DFT the transformed overall signal becomes

$$X[n] = \frac{Ae^{j\phi}}{N}\sum_{k=0}^{N-1} e^{j2\pi(\frac{f}{fs}-\frac{n}{N})^k} + W[n], \quad n = 0, 1, 2, ..., N-1, \tag{9}$$

where $W[n] = DFT\{w[k]\}$.

After DFT operations, we implement Maximum Bin Search (MBS) as the coarse search stage of FFT-based estimation. The principle of MBS is to search for a frequency bin that contains the most energy among all bins.

Firstly, by summing and squaring the geometric series for the pure sinusoid part in equation (9) we obtain

$$|S[n]^2| = \frac{A^2}{N^2}\left|\frac{e^{2\pi N\frac{f}{fs}}-1}{e^{2\pi N(\frac{f}{fs}-\frac{n}{N})}-1}\right|, \quad n = 0, 1, 2, ..., N-1. \tag{10}$$

It can be easily shown that $n = \frac{Nf}{fs}$ maximizes the above expression and this n gives the location of the bin that contains the most energy among the Fourier coefficients.

In addition, the concept of resolution and its improvement can be explained with the help of the zero-padding algorithm. Initially when the sampling process is finished, we have N number of samples directly from the signal. If we map these N samples to the frequency domain, we are observing the energy distribution of the spectrum for N times, which is identical to the number of samples. By padding s[k], the original signal, with N_z zeros, no information has been added to the samples but we increase the number of observations on the spectrum. And this is because the number of points after DFT has now become $N + N_z$. Therefore, we can express the original resolution of spectrum observations as

$$\Delta f = \frac{fs}{2N}, \tag{11}$$

while the improved resolution after zero-padding is

$$\Delta f_z = \frac{fs}{2(N+N_z)}. \tag{12}$$

4.5 Cramer-Rao Lower Bound

According to Rife in [5], Cramer-Rao bound (CRB) can be derived by taking the diagonal elements of the inverted Fisher information matrix for an unbiased MLE. When both the phase and the amplitude are unknown, the CRB can be expressed as a function of N.

$$CRB = \frac{3\sigma^2 fs^2}{2\pi^2 A^2 N(N^2-1)} \leq var(\hat{f} - f), \tag{13}$$

where $\hat{f}$ denotes the estimated frequency.

Cramer-Rao bound sets the theoretically lowest variance for ML estimations and is often utilized as the benchmark for judging the performance of estimators. It also stands for the estimation variance of an unbiased ML estimator when outlier does not occur.

Importantly for this thesis, the proposition of thresholds are mostly with respect to CRB. In Section 6 we will see that Knockaert [7] defines the threshold onset to be the diverging point of Barankin bound from CRB and Quinn [4] and Aboutanios [15] express the threshold as the point in SNR when the mean square estimation error is twice the magnitude of CRB.

4.6 Threshold Effect

It is observed that when SNR drops to a specific level, the ML estimation MSE surges and the curve deviates from CRB, deteriorating the reliability of the estimated results. This specific point is termed the signal-to-noise-ratio threshold that corresponds to the N value of the ML estimation. Compared to the gradient of CRB, the rise in estimation error is so rapid that the results quickly become unreliable after the threshold [7]. On the other hand, just because this fast growth of MSE, estimates conducted at an SNR slightly higher than the threshold would still be accurate. This phenomenon calls for an accurate indicator of threshold onset and thereby motivates the following research on threshold behavior.

Apart from the deviation of MSE from CRB, we can also interpret threshold effect from the angle of the tighter lower bounds' behavior. For an achievable lower bound, its value at each SNR is defined to be the infimum of the errors of all estimators. That is to say, an achievable lower bound embodies threshold effect as SNR decreases.

The CRB theoretically sets the lowest attainable error for estimations. However, it is not achievable at low SNRs and cannot give any insight of threshold effect. In contrast, Barankin bound (BB) and Ziv-Zakai bound (ZB) show tigher track of the estimation error. Therefore, the departure of Barankin bound and Ziv-Zakai bound from CRB can roughly locate the threshold onset [10].

An example of unbiased ML estimation MSE is displayed in Figure 1. Compared with the blue line, which is the CRB, the MSE suddenly surges when SNR drops below -5dB and, within the next 2dB, deteriorates to higher than 40dB. This is equivalently illustrating that the average deviation of estimated frequency from the true frequency will suddenly increase to around 100Hz and this happens with an SNR just 2dB lower than the estimates that were still accurate.

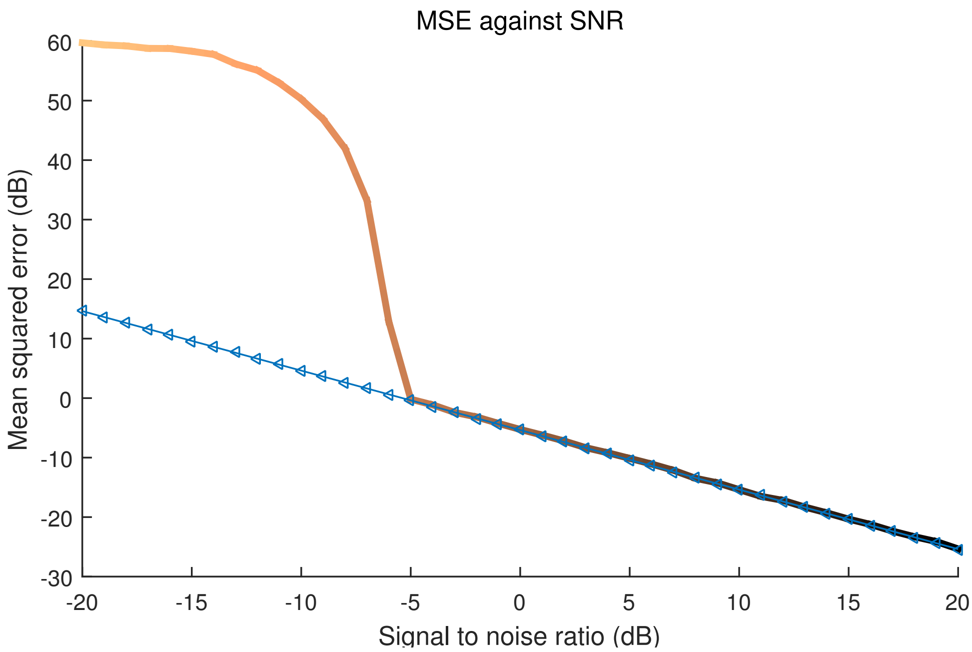

Figure 1: Threshold effect

5 Theoretical MSE and Outlier Probability

In this section, we aim to form basic understanding on how threshold effect is formed and what relationship the estimation error of MLE has with outlier probabilities. In order to theoretically analyze threshold effect, we introduce Rife and Boorstyn's formulation of theoretical mean square error of unbiased ML estimations [5]. Firstly, we define the outliers to occur when noise frequency bins are greater than the true frequency bin (i.e. MBS is intervened by heavy noise and fails to track the correct bin). Then, the formulation of the overall mean square error takes the weighted probability of estimation error of the outlier and no-outlier cases. When there is no outlier the mean estimation variance is approximated by CRB as in equation (13). In the case of outliers, the probability of selecting each noise bin follows *a priori* probability since distributions of the white noise bins are identical. The variance is thus given by the uniformly distributed probability spanning from $\frac{-fs}{2}$ to $\frac{+fs}{2}$, which is $\frac{fs^2}{12}$. By taking the weighted variance, the authors gave the following expression

$$MSE = q\frac{fs^2}{12} + (1-q)\frac{3fs^2}{2\rho\pi^2 N(N^2-1)}, \tag{14}$$

where q denotes the outlier probability. Looking again at Figure 1, we can now match the left half part of MSE (left to the threshold) with the left half part of equation (14) which is the proportion of variance attributed to outliers. And the same matching can be done for the right half part. This is because when SNR is low, q becomes large whereas $1-q$ is very small and vice versa.

Rife and Boorstyn started the formulation of the outlier probability by defining a DFT operator

A as

$$A_x = A(\frac{2\pi x fs}{T}) = \frac{1}{N}\sum_{n=0}^{N-1} Z_n exp(-\frac{j2\pi nx}{N}), \quad x = 0, 1, 2, ..., N-1, \tag{15}$$

where x denotes the index of frequency bin, A_x is the result of DFT operation and the vector Z_n is the subject being Fourier transformed, which in this case is our signal model $X[k]$.

We take the modulus of transformed signal and set up the variables for the two types of bin number:

$$\begin{cases} C_m = |A_m| & x = m \\ C_k = |A_k| & x \neq m, \end{cases} \tag{16}$$

where m is the index of true frequency bin.

An assumption is made for simplicity that the signal frequency is half the sampling frequency

$$f = \frac{fs}{2} \;\; so\ that \;\; m = \frac{N}{2}. \tag{17}$$

Based on this assumption and the presence of AWGN, C_k follows Rayleigh distribution since it is a positively valued random variable. C_m is Rice distributed which is centered at the Fourier coefficient of the original pure sinusoid. Thus, the probability density functions are

$$C_k = f_k(x) = \frac{Nx}{\sigma^2} exp(-\frac{Nx^2}{2\sigma^2}) \tag{18}$$

$$C_m = f_m(y) = \frac{Ny}{\sigma^2} exp(-\frac{N(y^2 + A^2)}{2\sigma^2}) I_0(\frac{NAy}{\sigma^2}), \tag{19}$$

where I_0 stands for the modified Bessel function of the first kind (BesselI).

It is worth mentioning that in this case the outlier probability refers to the total outlier probability, Q_N which stands for probability of all possible numbers of outliers. It is simpler to start by calculating 1 - Q_N, the probability that maximum bin contains the true frequency (i.e. no outliers).

$$\begin{aligned} 1 - Q_N &= p(all\ C_k < C_m) \\ &= \int_x P(all\ C_k < C_m | C_m = x) P(C_m = x) dx. \end{aligned} \tag{20}$$

Knowing that choosing each noise bin follows *a priori* probability, the first term $P(all\ C_k < C_m | C_m = x)$ can be expressed by

$$[P(C_1 < C_m | C_m = x)]^{N-1}. \tag{21}$$

Thus equation (20) becomes the integral of pdf functions

$$1 - Q_N = \int_0^\infty f_m(x) [\int_0^x f_k(y) dy]^{N-1} dx, \tag{22}$$

where the integral inside the bracket on the RHS is the Rayleigh-distributed probability of bins when $x \neq m$, which is

$$\begin{aligned} \int_0^x f_k(y) &= \int_0^x \frac{Ny}{\sigma^2} exp(-\frac{Ny^2}{2\sigma^2}) dy \\ &= 1 - exp(-\frac{Nx^2}{2\sigma^2}). \end{aligned} \tag{23}$$

Substituting equation (23) into equation (22), we have

$$1 - Q_N = \int_0^{\infty} (1 - exp(-\frac{Nx^2}{2\sigma^2}))^{N-1} \cdot \frac{Nx}{\sigma^2} exp(-\frac{N(x^2 + A^2)}{2\sigma^2}) I_0(\frac{Ax}{\sigma^2}) dx. \tag{24}$$

The authors simplified the above equation to the form of sums of factorials and exponents. However, the implementation of the new expression transcends our programmable upper limit of numbers and also the precision of handling small fractions. It is more applicable to use equation (24) to numerically calculate outlier probability with computer programs. The result is shown in Figure 2.

After this, we substitute the result of outlier probability to equation (14) and the theoretical mean square error of ML estimation is calculated as displayed in Figure 3. This theoretical analysis of unbiased ML estimation error facilitates our study of several threshold expressions in the next sections.

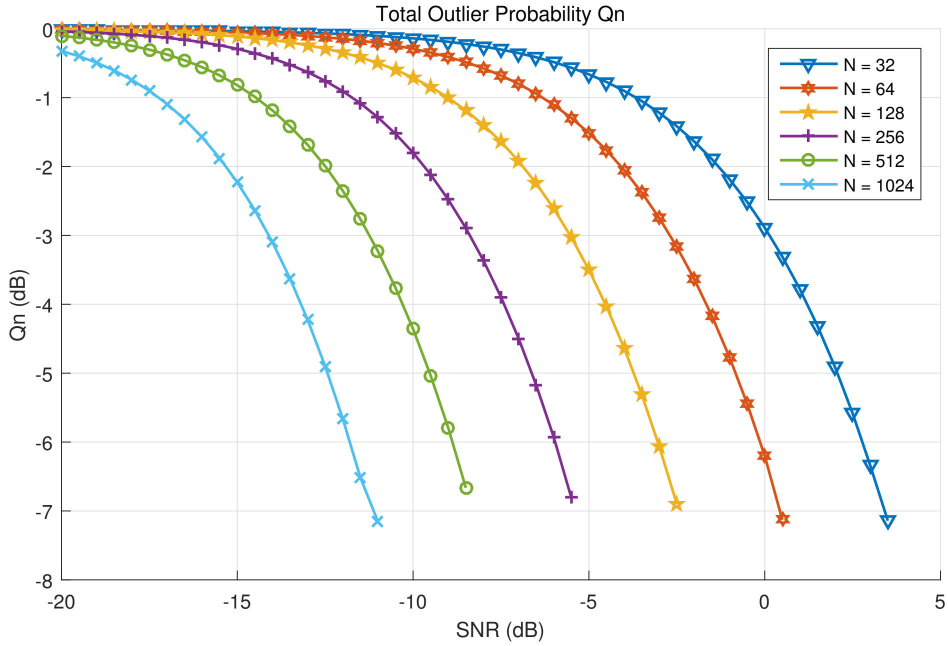

Figure 2: Outlier probability

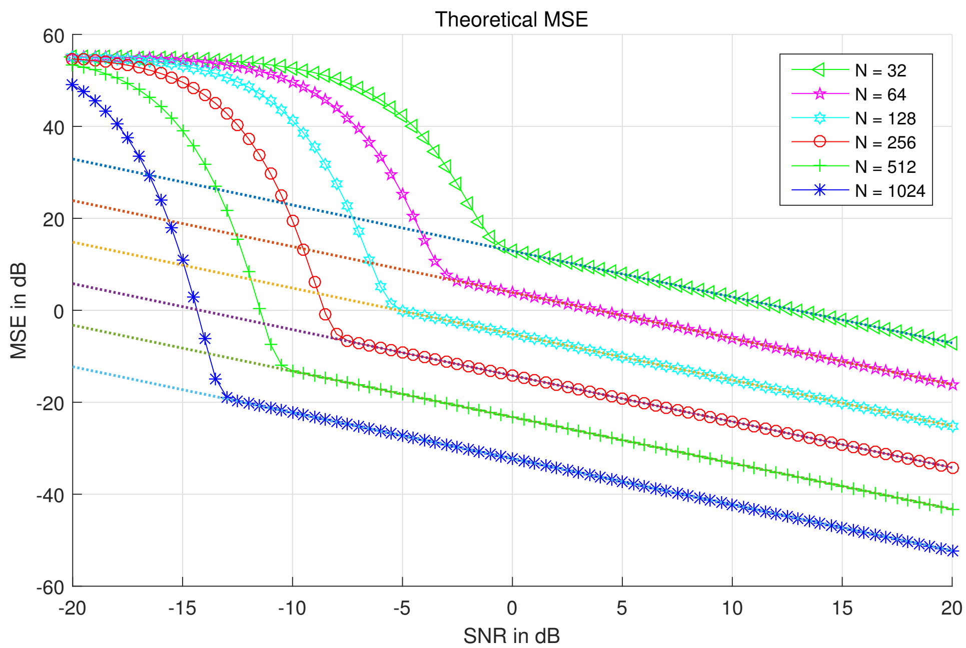

Figure 3: Theoretical MSE

6 Threshold Expressions in the Literature

In this section, we study and compare three interpretations of the threshold for ML estimation error. The methods differ in both their methods of modelling the thresholds and the resulting threshold values. We criticise on the threshold expressions by their accuracy in locating the turning points of MSE (i.e. the 'knees' of MSE). Nevertheless, this is just a general criterion and for specific utilities and conditions other criteria can be included. This is exemplified at the end of Section 6.3 where we introduce an extra term to scale Q&A threshold for small Ns to level a tolerable amount of estimation error.

We examine each expression by calculating the thresholds and plotting them on the theoretical MSE. As mentioned before, although the outlier probability in equation (24) cannot be easily simplified and solved for an expression of SNR, it can be computed by programs and substituted to the MSE equation. We use this theoretical MSE for Ns of 32, 64, 128, 256, 512 and 1024 as a basis to visually present the thresholds and make comparison among them.

6.1 Williamson, Anderson and Brian's Threshold Expression

Williamson, Anderson and Brian formulated a threshold expression by firstly defining "a physically non-implementable" estimator whose error variances are identical to CRB in [8]. Then he choose a criterion for this estimator to turn from accurate to inaccurate and it is utilized as threshold indicator. This indicator parameter happen to have similar expression with the phase MSE of the estimator above, which is then determined by Phase Locked Loop algorithm to be roughly constant at 0.0625 rad^2. This method is declared to be also viable for real case signals (sine functions). Brief procedure of derivation is given below.

Recall that the ML estimated frequency is given by

$$\hat{f} = argmax|A(f)|$$

Assume that the N-sampled signal model in equation (1) is used in this paper. The authors normalized the estimation error by $\frac{fs}{2}$ for simplicity such that

$$\delta = (\hat{f} - f)\frac{1}{2fs}. \tag{25}$$

Then they introduced a method of approximating the main lobe of the spectrum by

$$|A(\delta)|^2 \approx \alpha_0 + \alpha_1\delta + \alpha_2\delta^2 \tag{26}$$

, where A is the DFT operator, α_0 α_1 and α_2 are constants. Then the estimated estimation error is improved to

$$\delta^* = \frac{-\alpha_1}{2\alpha_2} \tag{27}$$

The estimation can therefore be done by finding the parameters that locates the maximum point on the approximated curve and is said to be accurate when the indicator $N\delta$ is small. By implementing this algorithm and calculating parameters including α_0 α_1 and α_2, the MSE of the approximated normalized estimation frequency is expressed as follows.

$$E(\delta^*)^2 = \frac{3\sigma^2}{A^2 N^3}, \tag{28}$$

and through similar process the phase MSE is given by

$$E(\hat{\phi} - \phi_0)^2 = \frac{4\sigma^2}{A^2 N^3}, \tag{29}$$

where E denotes mathematical expectation. The threshold sits where the estimation is just going to become inaccurate. And as stated above, this happens when the indicator $N\delta$ becomes large so that the indicator for threshold onset can be chosen as its MSE, which is transformed from equation (28).

$$E(N\delta^*)^2 = \frac{3\sigma^2}{A^2 N} \tag{30}$$

We notice that this threshold indicator is actually $\frac{3}{4} \cdot phaseMSE$. We substitute the definition of signal to noise ratio for single tone in equation (3) and have

$$E(N\delta^*)^2 = \frac{3}{\rho N} \tag{31}$$

In the authors' experiments they used PLL to study the phase MSE and found that it is roughly constant at around 0.0625 rad^2. By substitution we can write the final threshold expression as

$$SNR \approx 10 * log(\frac{3}{0.0625N}). \tag{32}$$

The results are shown in Table 1 and Figure 4.

Table 1: Williamson et al's Threshold Expression

N	Threshold SNR (dB)
32	1.7609
64	-1.2494
128	-4.2597
256	-7.2700
512	-10.2803
1024	-13.29059

As can be seen from the plot, this threshold expression can roughly tell the location of threshold onset and keeps close track of the simulated 3dB turning point with large Ns. For example, for N of 2048 and 4096, the actual 3dB turning point is around 16.3 dB and 18.7 dB. And Williamson et al's threshold is at 16.03 dB amd 19.0 dB which are close to the theoretical turning points. However, this indicator cannot accurately locate the 'knees' with lower Ns which are more often used in practice. Just as the authors stated in the paper, this method is not proposed from the angle of outlier probabilities so we do not use this as a reference for the following sections.

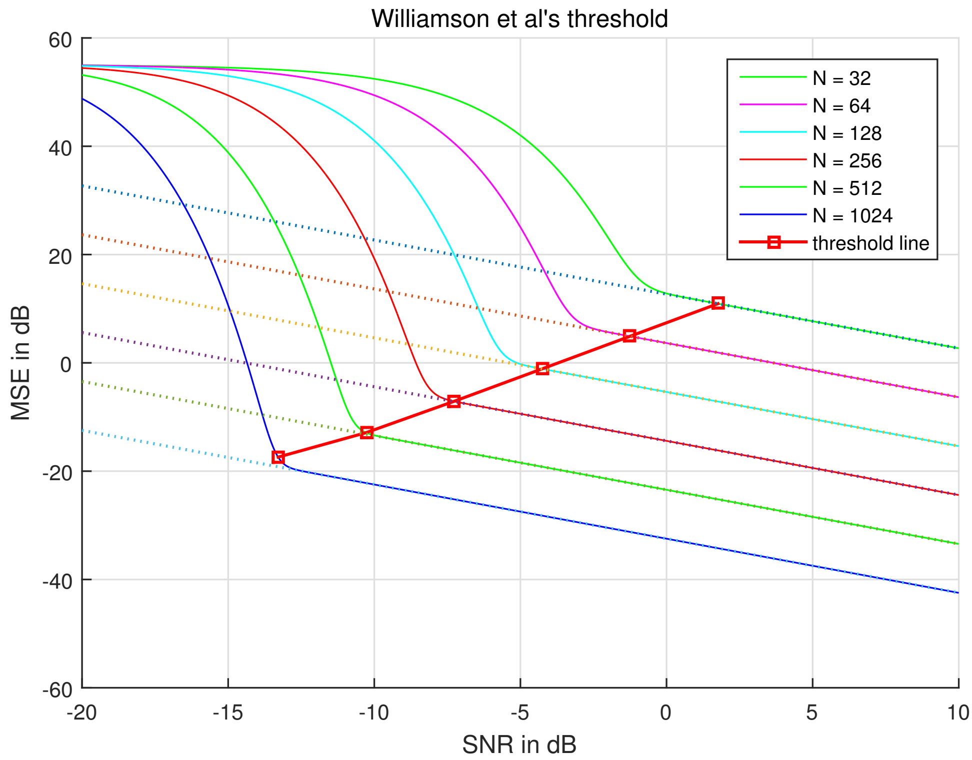

Figure 4: Williamson et al's threshold

6.2 Knckaert's Threshold Indicator

In 1997, Knockaert proposed an approach of defining the threshold in terms of the ratio between the tighter lower bounds and the CRB [7]. As has been roughly introduced previously, CRB is theoretically the lowest possible estimation variance for unbiased ML estimation and is only achievable at high SNRs. In other words, CRB provides no insight of threshold effect but the tighter lower bounds, such as the Barankin bound (BB) and Ziv-Zakai bound (ZB), are able to outline the threshold effect. Therefore, one could reasonably define the threshold to occur when BB or ZB deviates from CRB.

Even though it is an inspiring approach to conceptualize threshold effect via the ratio between lower bounds, Knockaert's this expression is shown to be inaccurate. Thus, we will keep brief by looking at the threshold indicator and observe the thresholds on a plot. The author stated that when N, the number of samples is large enough, threshold is expected to occur when

$$\rho \times \frac{N}{ln(N)} < 1 \tag{33}$$

Because of the different definitions of signal to noise ratio in the author's paper and in this thesis, we modify the expression using our definition in equation (3) and the indicator expression becomes

$$\rho < \frac{2ln(N)}{N} \tag{34}$$

We work out the resulting values for the thresholds and present them in Table 2 and Figure 5. As the author concluded in the paper, the result may be plagued with the biasedness in the Barankin bound and cannot track the knees of the mean square error curve. We also observe that although the lower bounds are shown to embody threshold effect, currently there is no one lower bound that is tight enough to accurately locate the onset of threshold effect. Some of the optimal bounds are described in [16] and [17] and in these papers the authors also give comparisons of the actual MSE against a variety of existing lower bounds. From their illustrations, one can obviously point out that the 'threshold' in classical Barankin Bound is shifted to smaller SNRs with significant difference with threshold regions of actual MSE in shape. This can be one of the reasons why the ratio of Barankin Bound against CRB is not an optimal threshold indicator. Further research may outline tighter lower bounds and may be provide better threshold indicator using their behaviours.

Table 2: Knockaert's Threshold Indicator

N	Threshold SNR (dB)
32	-6.642
64	-8.861
128	-11.20
256	-13.63
512	-16.13
1024	-18.68

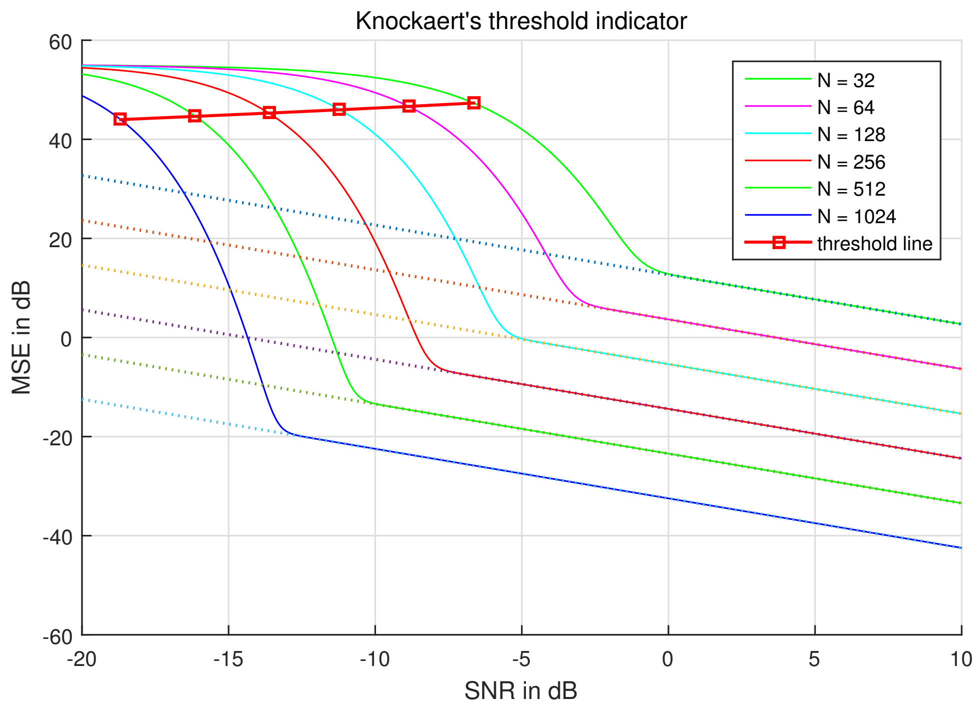

Figure 5: Knockaert's threshold

6.3 Q&A Threshold Expression

In 1994, Quinn and Kootsoskos [4] gave a threshold expression that aimed to indicate the point of SNR at which the MSE of unbiased MLE with the signal model in equation (1) becomes twice as large as CRB. That is,

$$MSE - CRB = 3dB$$

In 2002, Aboutanios [15], detected a mistake in their calculation of the outlier probability and corrected the mistake to give an improved threshold expression. This expression is reported to be very accurate and we could consequently consider ML estimations with lower SNRs than the thresholds to be unreliable. We refer to this threshold expression as the Q&A threshold in later Sections and use it as a benchmark for investigating outlier probabilities and the new threshold expression. This section will briefly cover the main concepts amid the derivations, the correction of the mistake and the analysis of the resulting threshold expressions.

Quinn and Kootsookos' derivation started with the normalized MSE expression w.r.t. CRB:

$$\frac{MSE}{CRB} - 1 = q\ (\frac{fs^2}{12}\frac{1}{CRB} - 1), \tag{35}$$

where we recall that CRB is

$$CRB = \frac{3fs^2}{2\pi^2 \rho N(N^2-1)} \leq var(\hat{f} - f).$$

The authors chose a different way of calculating Q_N from that of Rife and Boorstyn in [5] and managed to arrive at partitioned convergent integrations that were easier to manage. The procedure starts with modelling total outlier probability, Q_N.

$$Q_N = P\{\max_{n \neq m} |x(n)^2| > |X(m)^2|\}. \tag{36}$$

For the sake of simplicity, we again assume that the maximum bin $m = \frac{N}{2}$ (i.e. f $= \frac{fs}{2}$) such that

$$Q_N = P\{\max_{n \neq \frac{N}{2}} |x(n)^2| > |X(\frac{N}{2})^2|\}. \tag{37}$$

Recall that White Gaussian Noise has a mean of zero and a variance of sigma and we utilize this property to simplify our derivation by denormalizing X(n) to Z(n) where

$$Z(n) = \frac{2N|X(n)|^2}{\sigma^2}. \tag{38}$$

Note that X(n) is the DFT of the noisy signal and we can evaluate $Z(\frac{N}{2})$ by

$$Z(\frac{N}{2}) = \frac{2N}{\sigma^2}\left|\frac{1}{N}\sum_{k=0}^{N-1}(Ae^{jk2\pi\frac{f}{fs}} + w(k))e^{-jk2\pi\frac{n}{N}}\right|^2. \tag{39}$$

Noting that $\frac{f}{fs} = \frac{n}{N}$, we have

$$\begin{aligned}
Z(\frac{N}{2}) &= \frac{2N}{\sigma^2}\left|\frac{1}{N}\sum_{k=0}^{N-1}(Ae^{jk\pi} + w(k))e^{-jk\pi}\right|^2 \\
&= \left|\frac{\sqrt{2}}{\sqrt{N}\sigma}\sum_{k=0}^{N-1}(A + w(k)e^{-jk\pi})\right|^2 \\
&= \left|\sqrt{\frac{2}{N}}\sum_{k=0}^{N-1}(-1)^k\frac{w(k)}{\sigma} + \sqrt{2N}\frac{A}{\sigma}\right|^2 \\
&= (U + \sqrt{2N\rho})^2 + V^2,
\end{aligned} \tag{40}$$

where U and V are the real and imaginary parts of

$$\sqrt{\frac{2}{N}}\sum_{k=0}^{N-1}(-1)^k\frac{w(k)}{\sigma}. \tag{41}$$

Then Q_N is the probability that at least one Fourier coefficient other than that of the maximum bin is larger than the maximum-bin coefficient. Expressing in terms of Z we have

$$\begin{aligned}
Q_N &= p(\hat{Z} > Z(\frac{N}{2})) \\
&= P(\hat{Z} > (U + \sqrt{2N\rho})^2 + V^2) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{1 - [1 - e^{-\frac{1}{2}(x^2+y^2)}]^{N-1}\}\Phi(x)\Phi(y - \sqrt{2N\rho})dxdy,
\end{aligned} \tag{42}$$

where $\hat{Z}$ is the maximum of all squared Fourier coefficients and $\Phi(x)$ is the standard normal density function, given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \tag{43}$$

Equation (42) is an equivalent expression to one minus equation (24) where the integrations became difficult to simplify to an operable form. The authors solved the problem by partitioning the integral in a way such that each part converges to an operable expression. This was motivated by the Rayleigh distributed format of Fourier coefficients for each noise bin, which has been given in equation (18) as

$$f_k(c_k) = \frac{NC_k}{\sigma^2}exp(-\frac{NC_k^2}{2\sigma^2}).$$

After noticing that Ck − 2log(N) converges in distribution, the authors set

$$\frac{NC_k}{\sigma^2} = x + 2\ln(N), \tag{44}$$

such that

$$\begin{aligned}
P(\hat{Z} - 2\ln(N) \le x) &= [1 - e^{-\frac{1}{2}(x+2\ln N)}]^N \\
&= (1 - \frac{1}{N}e^{-\frac{1}{2}x})^N \to e^{e^{-\frac{1}{2}x}}.
\end{aligned} \tag{45}$$

Therefore the integral is partitioned into 3 sections ω_1, ω_2 and ω_3 with respect to the x values ranged by the sets $(-\infty, -\sqrt{2\ln N})$, $(-\sqrt{2\ln N}, \sqrt{2\ln N})$ and $(\sqrt{2\ln N}, \infty)$.

However, a mistake is made when the authors considered the term $(1-\frac{1}{N}e^{-\frac{1}{2}x})^N$ converges to $e^{e^{-\frac{1}{2}x}}$ from above so that

$$\int_{-\infty}^{\infty}\{1-[1-e^{-\frac{1}{2}(x^2+y^2)}]^{N-1}\}\Phi(x)dx < \int_{-\infty}^{\infty}(1-e^{-e^{-\frac{1}{2}x^2}})\Phi(x)dx \approx 0.4860709. \tag{46}$$

Aboutanios detects this mistake and found that the convergence is actually from below and that

$$(1-\frac{e^{-\frac{x^2}{2}}}{N-1})^{N-1} \leq e^{-e^{-\frac{x^2}{2}}}, \tag{47}$$

$$\{1-[1-\frac{e^{-\frac{x^2}{2}}}{N-1}]^{N-1}\}\Phi(x) \geq (1-e^{-e^{-\frac{1}{2}x^2}})\Phi(x). \tag{48}$$

One can show that there is no value for of K above which the inequality is satisfied. Therefore, another approach is utilized to partition the integral. A variable $\mu = -\sqrt{2ln(N-1)} - \epsilon$ is set up for some $\epsilon > 0$. Sustituting into the LHS of equation (45),

$$[1-\frac{e^{-\frac{x^2}{2}}}{N-1}]^{N-1} = \left(1-\frac{1}{k}e^{-\frac{\epsilon^2+2\epsilon\sqrt{2ln(N-1)}+x^2}{2}}\right)^{N-1}. \tag{49}$$

We then proceed to find some values of ϵ for the inequality to stand and hold for all values of x

$$\left(1-\frac{1}{k}e^{-\frac{\epsilon^2+2\epsilon\sqrt{2ln(N-1)}+x^2}{2}}\right)^{N-1} > e^{-e^{-\frac{x^2}{2}}}. \tag{50}$$

Due to the monotonicity of exponential function, the inequality is satisfied by some ϵ that hold the logarithm of the above expression which gives

$$f(x) = (N-1)ln\left(1-\frac{1}{k}e^{-\frac{\epsilon^2+2\epsilon\sqrt{2ln(N-1)}+x^2}{2}}\right) + e^{-\frac{x^2}{2}} > 0. \tag{51}$$

By differentiating f(x), we find a pair of solutions that give a local maximum at x = 0 with $f(0)^{'} = 0$ when the f(x) curve is concaving down (i.e. $f(0)^{''} < 0$). The solution is expressed as

$$\epsilon_N = -\sqrt{2ln(N-1)} \pm \sqrt{2ln(N-1) - 2ln[(N-1)(1-e^{-\frac{1}{N-1}})]}.$$

Since ϵ is defined to be positive, the only solution is thus

$$\epsilon_N = -\sqrt{2ln(N-1)} + \sqrt{2ln(N-1) - 2ln[(N-1)(1-e^{-\frac{1}{N-1}})]} \tag{52}$$

Substituting to include ϵ in our partitioning process, $\mu_N = \sqrt{2}ln(N-1) + \epsilon_N$ and the integral is separated in five sections: $(-\infty, -\mu_N)$, $(-\mu_N, -\sqrt{2\ln(N-1)})$, $(-\sqrt{2\ln(N-1)}, \sqrt{2\ln(N-1)})$, $(\sqrt{2\ln(N-1)}, \mu_N)$, and (μ_N, ∞).

After completing the integration and substituting outlier probability q back to equation (14), the author arrived at a threshold expression of

$$SNR_{Q\&A}(dB) = 10\log_{10}\{\frac{1}{N}[6\ln(N-1) + 2\ln(\ln(N-1)) + 4\ln(\pi) - 2\ln(6) + 2\epsilon_N^2] \tag{53}$$

. The calculated thresholds are given in Table 3 for several N-values. A visual presentation is given in Figure 6.

Table 3: Q&A Threshold Expression

N	Threshold SNR (dB)
32	-1.2373
64	-3.4834
128	-5.8586
256	-8.3227
512	-10.8525
1024	-13.4331

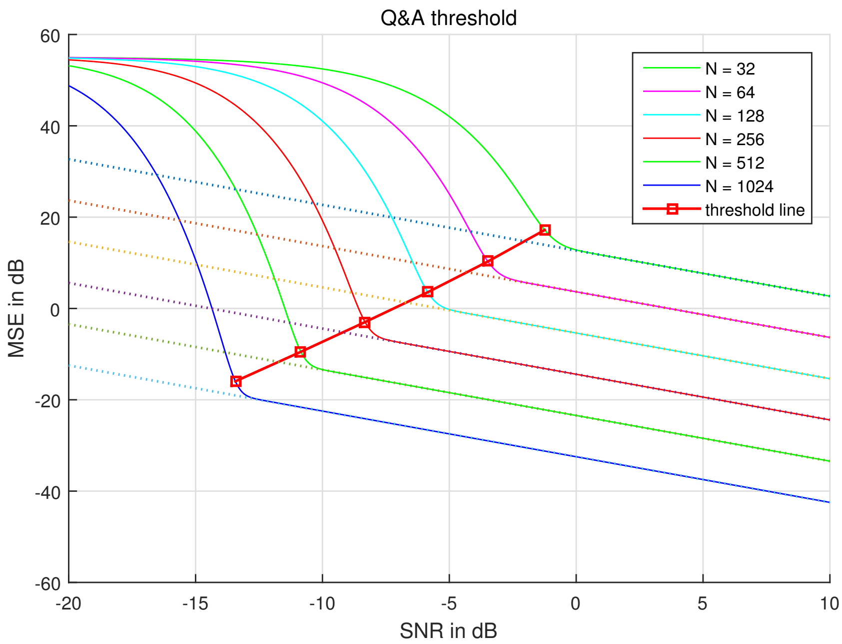

Figure 6: Q&A threshold indicator

As can bee seen from the plot, Aboutanios' threshold indicator is precisely located at the turning points of MSE curves. Meanwhile, the proposed definition of threshold, which is 3dB higher than the CRB, is approximately achieved. By keeping close track of the 'knees' of MSE, we can confidently avoid the severe deterioration of estimation error brought by the threshold effect. Also, estimates that are seemingly plagued with large noise level may be taken advantage of as long as their SNRs are above the threshold. This helps minimize waste of available signals.

However, even though the turning points are accurately pursued, we notice that with a N lower than 256, which is very commonly used in practice, the MSE is larger than 0dB even at the thresholds. That is to say, even at Q&A thresholds with 256 samples taken, we still suffer from 1Hz deviation on average for every single estimation. At an N of 32, MSE is around 20dB which infers an average estimation deviation of 10Hz. To avoid this substaintial error in estimation, we may have to sacrifice some SNR and warp the threshold by an additive term. For instance, an exponential term would serve the purpose of shifting the threshold largely with small Ns but very negligibly slightly when N is large. It could have similar form as equation (54) below. And the result would become that in Table 4 and Figure 7. Again, Q&A threshold will be used as the benchmark for the following research sections. However, we do not include this extra scaling term in the benchmark because for our purpose of studying outlier probability and threshold behavior, turning points are more significant.

$$SNR_{th}(dB) = SNR_{Q\&A} + 10e^{\frac{-(N-31)}{100}} \tag{54}$$

Table 4: Scaled Q&A Threshold Expression

N	Threshold SNR (dB)
32	8.7600
64	3.7815
128	-2.0211
256	-7.2554
512	-10.8525
1024	-13.4295

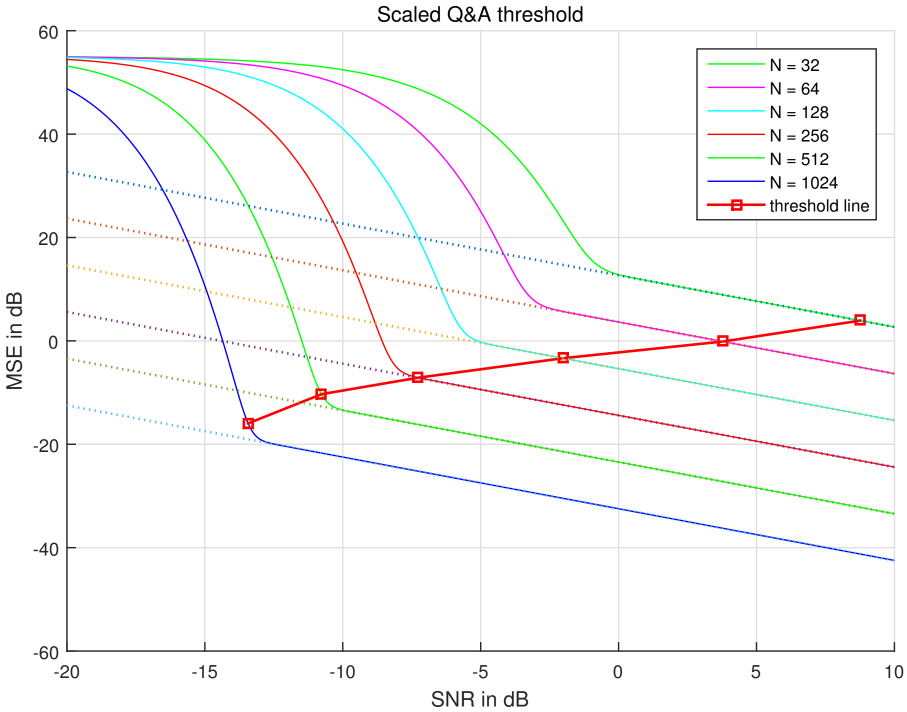

Figure 7: Scaled Q&A threshold indicator

7 Research on threshold behaviour

In this Section, we further analyse the threshold behaviour by focusing on the study of outlier probability, especially the single outlier probability. Detailed formulation and evidence will be given in support of propositions made on the probability of single outlier.

As mentioned in the definition of threshold effect, the MLE estimation error quickly jumps up when entering the threshold region, which turns the estimation result unreliable. This definition, however, remains vague and cannot provide clear insight of threshold effect. And it needs to be clarified which parameters affect the estimation MSE and how they dictate the properties of SNR thresholds. To probe deeply into the mechanism of threshold effect, we need to focus again on the composition of estimation MSE, which is the combination of the variances of uniformly distributed noise and the CRB. These two types of variances are weight by the probability of having and not having outliers (equation (14)).

$$MSE = q\frac{fs^2}{12} + (1-q)\cdot CRB \tag{55}$$

In the composition of MSE, fs is constant as the sampling frequency of estimation is fixed. CRB is also determined by equation (13). Thus, the only crucial variable in MSE is outlier probability q. Recall that q is the probability of the maximizer in Maximum Bin Search failing to locate the true frequency. In other words, it includes the probability of any possible number of noise frequency bins having a magnitude greater than the true-frequency bin on the Fast-Fourier-Transformed spectrum. And each slice of the probability, which is the chance of exactly some number of noise bins being higher than true frequency bin, is formed by the statistical model of choosing some number of the total N-1 Rayleigh distributed variable 'B's to be greater than one Rice distributed variable 'A' while the rest of 'B's are not.

It is noticed that when SNR is high, which infers a low likelihood (of the order 1E-5 to 1E-8) for outliers to occur and a high probability for it not to occur, the probability of one single outlier would be much larger compared with that of exactly two or more outliers even though the binomial coefficient would increase. This phenomenon motivates my further study on single outlier probability to clarify the outlier behaviour and thereby the threshold behaviour.

7.1 Universal Expression for Probability of an Exact Number of Outliers

To better examine and illustrate the findings in later this section, we formulate the statistical model for the probability of exactly L number of outliers. The method of derivation is inspired by that for total outliers written by Rife and Boorstyn (1974) mentioned in Section 5.

Suppose in the following system, we have (N-1) B random variables and a single A random variable, where B refers to the magnitude of any one other bin except than true frequency bin and A is the magnitude of the true frequency bin. We know that A follows Rician distribution and B follows Rayleigh distribution.

We introduce the following process To find P(some Bs > A). For all sizes of A, we use the integration w.r.t. A:

$$P(someBs > A) = \int_0^{\infty} P(someBs > A)dA. \tag{56}$$

Let the number of Bs greater than A be L, so

P(L number of B > A)

$$= \binom{N-1}{L} \int_0^{\infty} P(LnumberofBs > A), N-1-LnumberofBs < A)dA,$$

$$where \begin{cases} P(LnumberofBs > A) = P(B > A|A = x)^L \cdot P(A = x) \\ P(N-1-LnumberofBs < A) = P(B < A|A = x)^{N-1-l} \cdot P(A = x), \end{cases}$$

$$where \;\; P(LnumberofBs > A) = P(B > A|A = x)^L \cdot P(A = x),$$

$$and \;\; P(N-1-LnumberofBs < A) = P(B < A|A = x)^{N-1-l} \cdot P(A = x),$$

and the integrand expression is the joint probability of the two events.

So P(L number of Bs > A)

$$\begin{aligned} &= \binom{N-1}{L} \int_0^{\infty} [P(B > A|A = x)]^L [P(B < A|A = x)]^{N-1-l} \cdot P(A = x)dx \\ &= \binom{N-1}{L} \int_0^{\infty} [P(B > A|A = x)]^L [1 - P(B > A|A = x)]^{N-1-l} \cdot P(A = x)dx. \end{aligned} \tag{57}$$

Since the two events are related only by the mutual dependence on A, their joint probability is their product times the probability of the mutual dependency. And choosing P(A=x) rather than P(B = x) is for simplicity since A is Rice distributed, which is more complicated and more difficult to handle at higher orders than B.

As can be seen in equation (57), in the threshold region when outliers are just about to occur, the left part is small and the right part is large. Transforming from single outlier probability to double or multiple outlier probability would employ larger orders of left part and lower order of right part and therefore will be reduced significantly. This is the reason for single outlier being overwhelmingly large among probabilities of all numbers of outliers. We now derive the universal expression for L Bs being greater than A by using the statistical models of corresponding events.

A and B can be expressed by the following two distributions mentioned in equation (18) and equation (19) respectively.

$$A = f_k(x) = \frac{Nx}{\sigma^2} exp(-\frac{Nx^2}{2\sigma^2})$$

$$B = f_m(y) = \frac{Ny}{\sigma^2} exp(-\frac{N(y^2 + A^2)}{2\sigma^2}) I_0(\frac{NAy}{\sigma^2}),$$

such that

$$
\begin{aligned}
P(B > A|A = x) &= \int_0^x f_k(\lambda)d\lambda \\
&= \int_0^x \frac{N\lambda}{\sigma^2} exp(-\frac{N\lambda^2}{2\sigma^2})d\lambda \\
&= 1 - exp(-\frac{Nx^2}{2\sigma^2}).
\end{aligned} \tag{58}
$$

We substitute the distributions and equation (58) into equation (57) and will arrive at

P(L number of Bs > A)

$$
= \binom{N-1}{L} \int_0^\infty (1 - exp(-\frac{Nx^2}{2\sigma^2}))^L \cdot exp(-\frac{Nx^2}{2\sigma^2})^{N-1-L} \cdot \frac{Nx}{\sigma^2} exp(-\frac{N(x^2 + A^2)}{2\sigma^2}) I_0(\frac{Ax}{\sigma^2})dx. \tag{59}
$$

7.2 Expression for Single Outlier Probability

Using the model in equation (57), we formulate probability for single outlier.

P(single B > A)

$$
\begin{aligned}
&= (N-1) \int_0^\infty [P(B > A|A = x)] \cdot [1 - P(B > A|A = x)]^{N-2} \cdot P(A = x)dx \\
&= (N-1) \int_0^\infty (1 - exp(-\frac{Nx^2}{2\sigma^2})) \cdot exp(-\frac{Nx^2}{2\sigma^2})^{N-2} \cdot \frac{Nx}{\sigma^2} exp(-\frac{N(x^2 + A^2)}{2\sigma^2}) I_0(\frac{Ax}{\sigma^2})dx.
\end{aligned} \tag{60}
$$

This expression provides an important approach for us to analyze threshold effect from the view of single outlier probabilities which then plays an overwhelmingly significant role in the total outliers probability in the threshold region.

7.3 $\frac{Q_1}{Q_N}$ Ratio behaviour

7.3.1 Theoretical result for $\frac{Q_1}{Q_N}$ ratio

The calculation of single outlier probability can be conducted with programs such as MATLAB or Maple, despite the fact that it is difficult to find the inverse function and solve for SNR.

Firstly the ratio of single outlier probability against total outliers probability is calculated by the ratio of equation (24) and equation (60). We take only the ratios at the Q&A thresholds in order to compare with experimental result in the following subsection and the result is given in Figure 8.

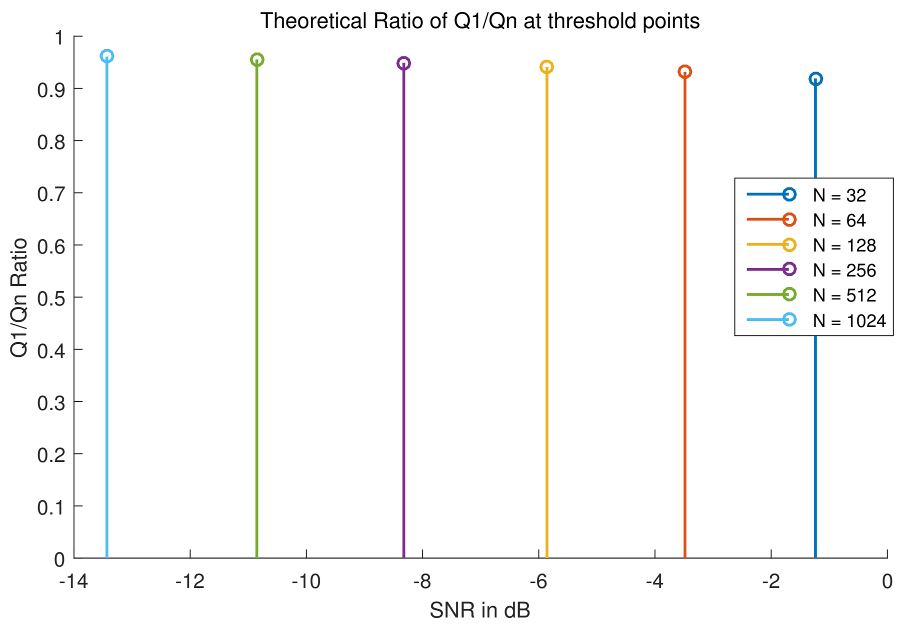

Figure 8: Theoretical ratio of $\frac{Q_1}{Q_N}$ at Q&A thresholds

7.3.2 Experimental Result for $\frac{Q_1}{Q_N}$ Ratio

The experiment is conducted firstly by using FFT to map the constructed signal model to the frequency domain into N number of frequency bins and record the magnitude of each Fourier coefficients. Recall that the signal model in equation (1) consists of a complex single tone and the AWGN. For each number of N and only at the Q&A threshold corresponding to the N-value, this process is repeated 100 million times to give an accurate estimation. A row vector is declared to record the number of occurrence (0 to 100 million) of exactly each number (1 to N-1) of outliers that have occurred. After running 100 million repeated experiments, the first element of the vector corresponds to the number of occurrence of single outliers while the sum of all elements stands for number of total outliers and the single outlier probability is given by the ratio of the first element and the sum. The result is shown in Figure 9.

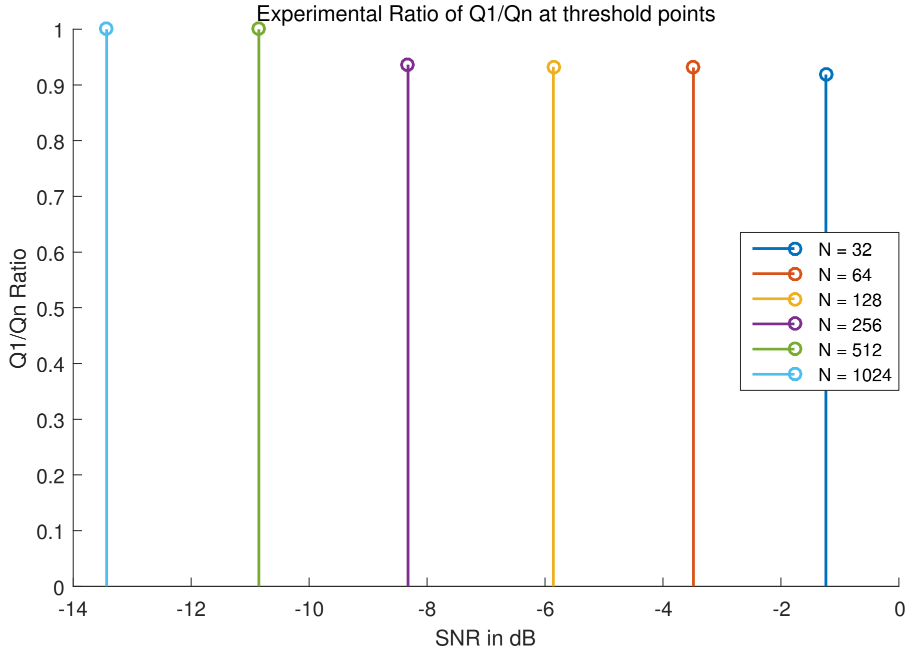

Figure 9: Experimental ratio of $\frac{Q_1}{Q_N}$

To compare the theoretical and experimental result, we plot the two in one graph below.

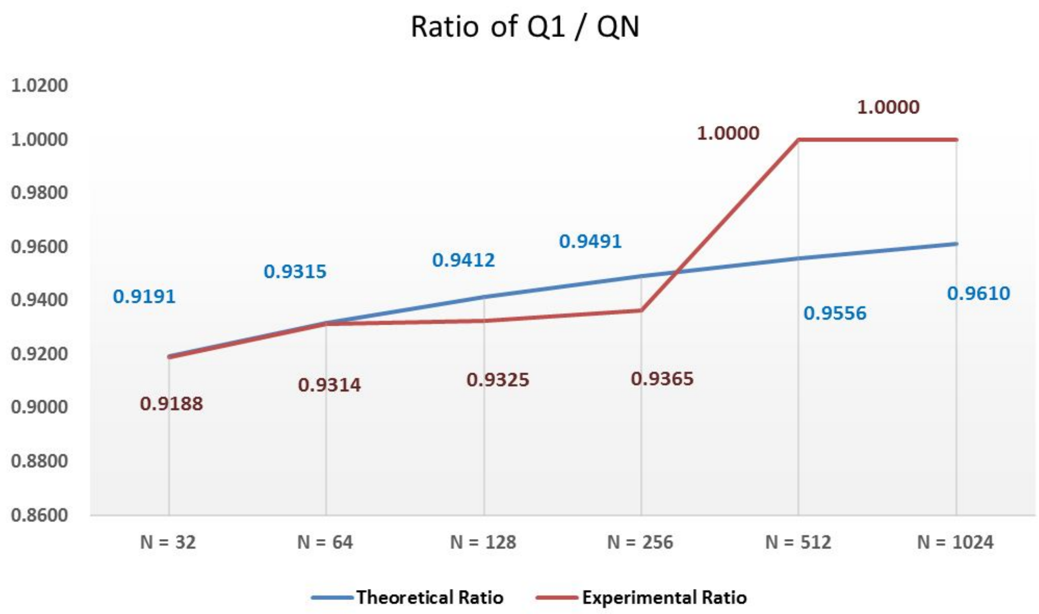

Figure 10: Theoretical and experimental ratio of $\frac{Q_1}{Q_N}$

Generally speaking, the proportion out of all occurrences of outliers taken by that of single outlier is at least 91.88% which is highly dominant. As N increases, this percentage also increases significantly to nearly unity. This is because the probability of having an outlier is lower when we have more information obtained from the sampling process. From single outlier to double outlier, the probability is multiplied by the probability of any one noise bin being greater than true frequency bin (as small as 1E-5 to 1E-8) and divides the probability of one minus it (nearly unity), despite the change in binomial coefficient (of order 1E3).

In experimental result, the average trend adheres to that of theoretical result, but in order to obtain a smoother curve one would have to repeat the experiment for one billion or more repetitions since for an event with the probability of P_x, it requires as many as $\frac{1-P_x}{P_x} \cdot \frac{1}{\epsilon}$ times experiments to produce a stable observation of the event. The ϵ is wanted the ratio of the estimates' standard deviation and the probability itself. The algorithm of deriving this is included in Appendix A.

7.3.3 Quantitative Behaviour of Q_1

The quantitative study of Q_1 is again based on the benchmark of Q&A threshold. We look at the magnitude of outlier probabilities and can precisely compare the results of theoretical and experimental single outlier probability as well as the total outliers probability. Data are recorded in Table 5 below.

Table 5: Quantitative analysis of Q_1

N	Q&A threshold (dB)	Theoretical Q_1	experimental Q_1	Theoretical Q_N
32	-1.2373	7.5559E-5	7.6291E-5	8.2212E-5
64	-3.4834	1.5589E-5	1.5610E-5	1.6736E-5
128	-5.8586	3.3551E-6	3.5900E-6	3.5646E-6
256	-8.3227	7.4092E-7	5.9000E-7	7.8065E-7
512	-10.8525	1.6649E-7	1.9418E-7	1.7423E-7
1024	-13.4331	3.7870E-8	7.7647E-8	3.9407E-8

Two examples of experimental and theoretical single outlier probability are given in the following figures for N-values of 64 and 128.

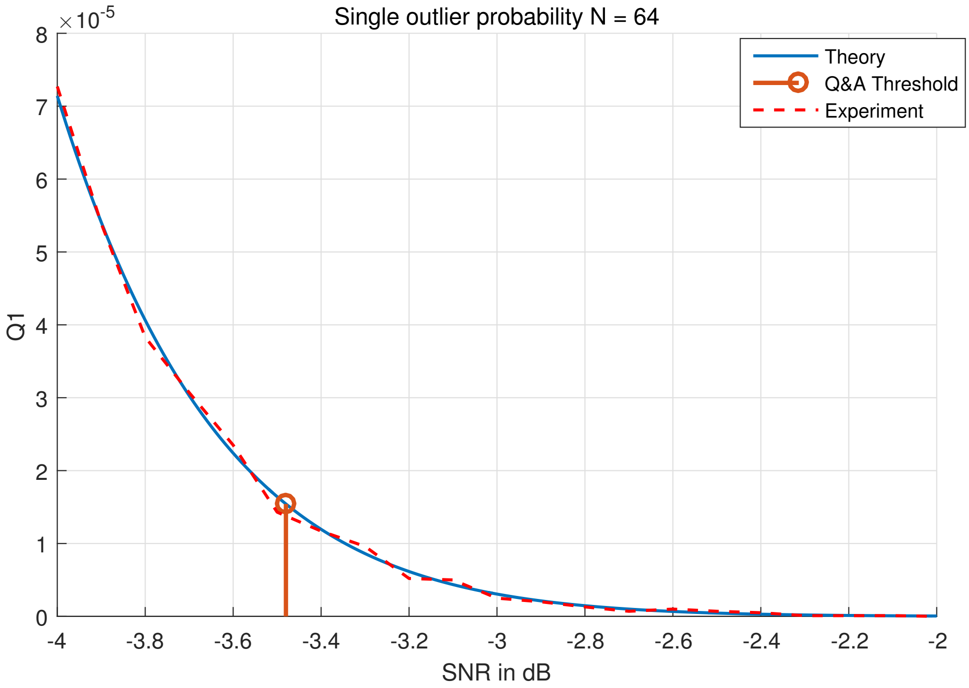

Figure 11: Experimental and theoretical Q_1 (N = 64)

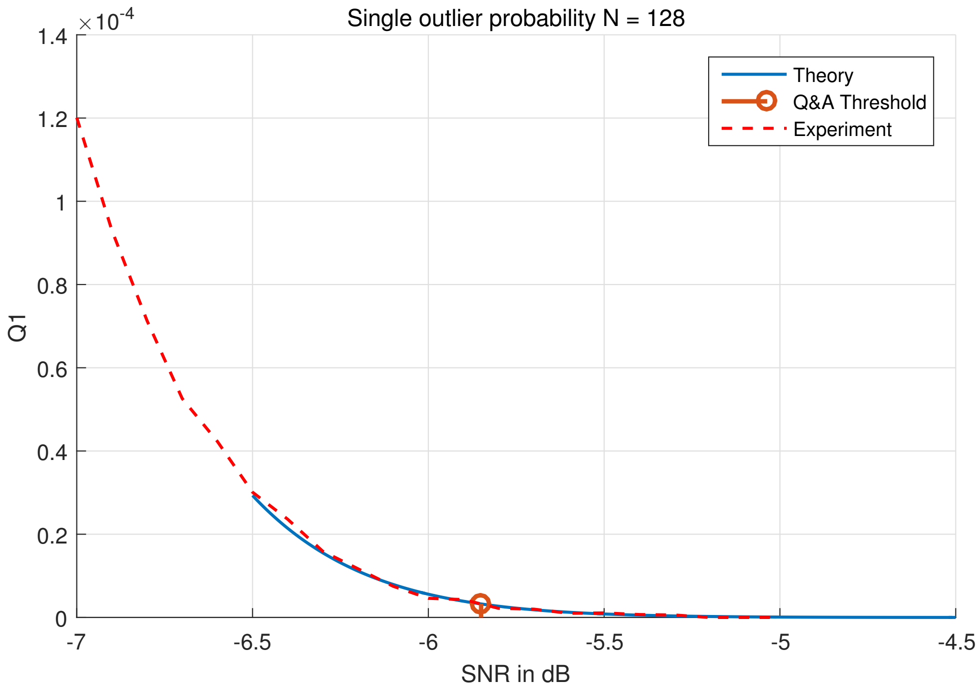

Figure 12: Experimental and theoretical Q_1 (N = 128)

The experiments were conducted with 10 million repetitions so that a probability of the order 1E-6 can be detected. It is shown that theoretical data is highly consistent with experimental result, which again proves the relationship between Q_1 and Q_N. Minor fluctuations can still be seen in the plots, which infers that larger number of repeated search is needed to produce smoother and more accurate curves.

7.3.4 Post-threshold Deviation

The phenomenon of single outlier being able to represent total outliers at large SNRs but not for all SNRs is visually presented in the following plot. Around the thresholds, both Q_N and Q_1 are very small but they share very similar magnitude as clarified in the previous subsections.

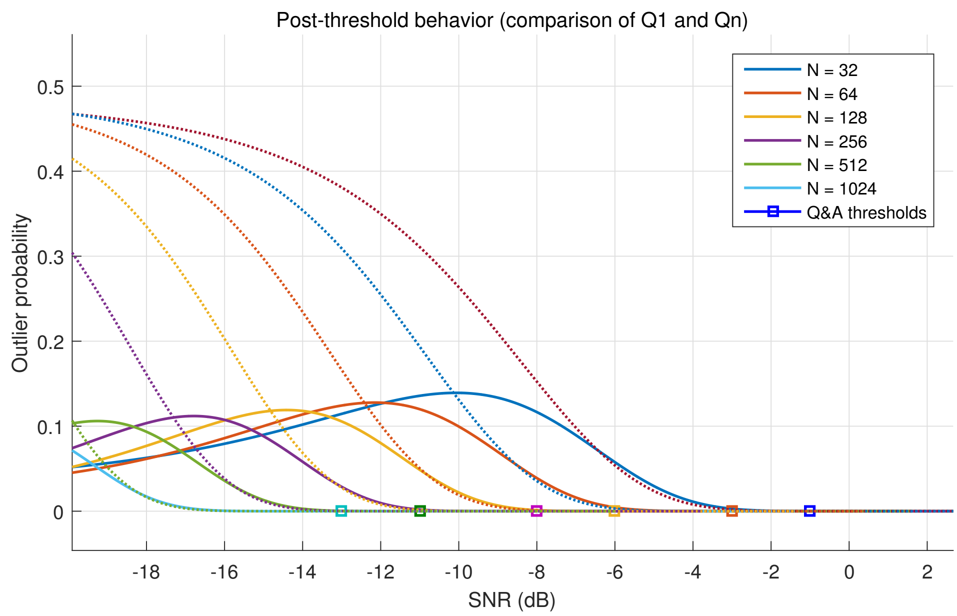

Figure 13: Post-threshold deviation of outlier probabilities

Also from the plots, single outlier probability deviates from total outliers probability as SNR deteriorates and will stop increasing at some point, and henceforth fall significantly when SNR continue to decrease. This is because at low SNRs, the chance of having an outlier becomes very large (up to around 0.5) so that double, triple and multiple outliers probabilities are comparable to Q_1. Due to the fact that estimates below the thresholds are usually discarded, we do not study further on the post-threshold deviation.

7.3.5 MSE Behaviour

A more illustrative way of showing single outlier probability's capability in representing total outliers in threshold regions is to substitute single outlier probability into the expression of MSE in equation (14). We visualize the comparison between Q_N and Q_1 by plotting the estimation MSE composed of Q_N and Q_1 respectively in one graph (Figure 14). The dashed lines represent estimation MSE that is formed by Q_1. They overlap with the curves formed by Q_N with SNRs greater than and around the Q&A threshold region. Q_1-MSE curves deviate from those of Q_N-

MSE only when SNRs are much lower than the Q&A thresholds. This indicates that in formulating thresholds for Maximum Likelihood Estimation based frequency estimators, the threshold can be formulated by single outlier probability instead of total outliers probability.

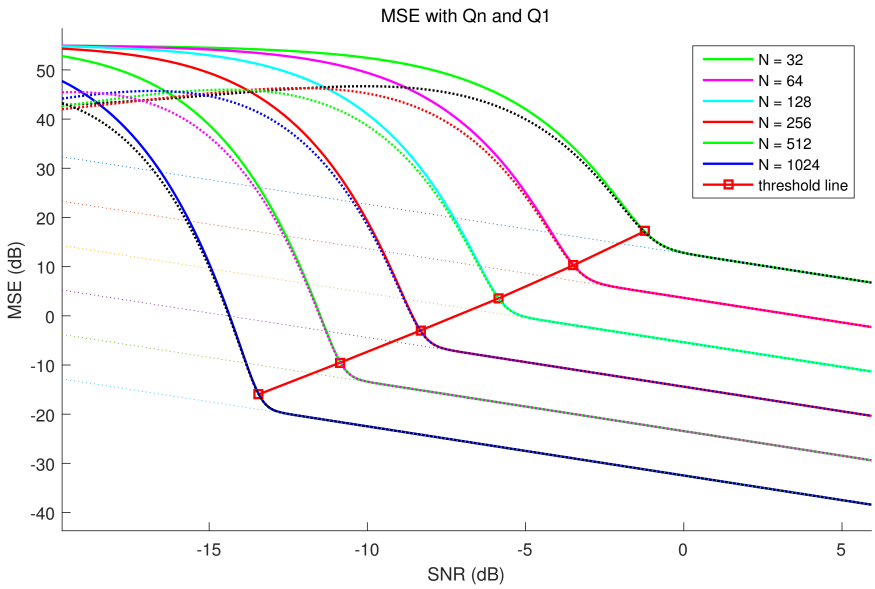

Figure 14: MSE behavior with Q_N and Q_1

In addition, this also infers that the single outlier probability can be used to express the threshold. With the properties being slightly different from that of total outliers probability, we can arrive at a different threshold expression using Q_1. The new expression is expected to be accurate since Q_1 is well capable of representing Q_N in the threshold region. The threshold formulating process can use either derivations or approximating techniques such curve-fitting and will be discussed in the next section.

8 Newly proposed threshold expression

Unlike the actual frequency estimations, thresholds usually only need to be determined once prior to estimations. They assist to distinguish accurate estimations from the inaccurate ones by setting the points of SNR lower than which estimation results should be rejected. The process of determining thresholds can be deemed as a type of calibration in which accuracy is the focus.

In the existing threshold expressions, the one proposed by Williamson et al could accurately locate the turning points for large Ns but not for Ns lower than 512. Knockaert's threshold is shown to be inaccurate. The expression by Quinn and Aboutanios can closely keep track of the 3dB turning point but assumptions during the derivation process may induce minor inaccuracies. Due to these reasons, a more accurate threshold expression can be proposed to more precisely locate the 'knees' of estimation MSE.

One method of expressing the threshold is to use equation (14) with substitution of the expressions of outlier probability. This probability can be the total outliers probability or single outlier probability since they have been proven to be very similar within the threshold region. Therefore equation (14) will become the following form.

Choosing 3dB deviation point in MSE from CRB,

$$\begin{aligned} MSE = 2CRB &= q\frac{fs^2}{12} + (1-q)\cdot CRB \\ (1+q)CRB &= q(fs^2/12), \end{aligned} \tag{61}$$

where the outlier probability q can be single outlier probability in equation (60) or total outliers probability in equation (24). One would need to substitute these outlier probabilities and solve for the relationship between SNR and N to obtain threshold expression.

However, the expressions of total and single outliers involve non-closed-form integrals containing the product of exponents and the Bessel function. Knowing that N is usually not sufficiently large for us to neglect the exponents, we would have to employ complicated ways of solving the non-closed-form integrals. Even if one manages to escape the integration, the lengthy expression cannot provide an easily operable threshold expression for actual implementations.

Nonetheless, there exists indirect ways of solving the integration such as the techniques used in the derivation of Q&A threshold expression. As mentioned in Section 6.3, the integral is partitioned in a number of sections so that the results would converge. Approximations and assumptions have to be made along the process that could possibly introduce minor accuracies.

Other methods, such as Laplace Asymptotic Method and Saddle Point Method, may possibly be used to simplify the integrands and degrade the outlier probability equations to closed-form integrals.

A newly proposed threshold expression is introduced via the rational-polynomial curve-fitting method in order to compare the curve-fitting results against the theoretically derived Q&A threshold. We use programs to compute the theoretical MSE using Q_1 and Q_N respectively and search for the 3dB bifurcation point of MSE and CRB. The approximation utilized 1009 entries of data as shown in Figure 19 as an illustration. The rational-polynomial method produces a fourth order relationship between threshold SNR and the number of samples (N) for the Q_N-MSE and the Q_1-MSE respectively. The results are displayed by equation (62), table (6), and Figures 15 and 16.

$$SNR = \frac{p_1 x^4 + p_2 x^3 + p_3 x^2 + p_4 x^1 + p_5}{q_1 x^3 + q_2 x^2 + q_3 x + q_4}. \tag{62}$$

Table 6: Threshold expression parameters

Q_N				Q_1			
p_1	-21.75	q_1	1489	p_1	-21.79	q_1	1512
p_2	-1.289E4	q_2	8.701E4	p_2	-1.323E4	q_2	9.131E4
p_3	-2.8E5	q_3	-6412	p_3	2.818E5	q_3	-1.001E5
p_4	1.695E4	q_4	-501.2	p_4	4.663E4	q_4	-4273
p_5	1086			p_5	4179		

And the threshold results are shown below:

Table 7: Threshold results by new expression

N	Q_N-MSE threshold (dB)	Q_1-MSE threshold
32	-1.1383	-1.1804
64	-3.4019	-3.4282
128	-5.8725	-5.8855
256	-8.3289	-8.3398
512	-10.8153	-10.8249
1024	-13.4292	-13.4349

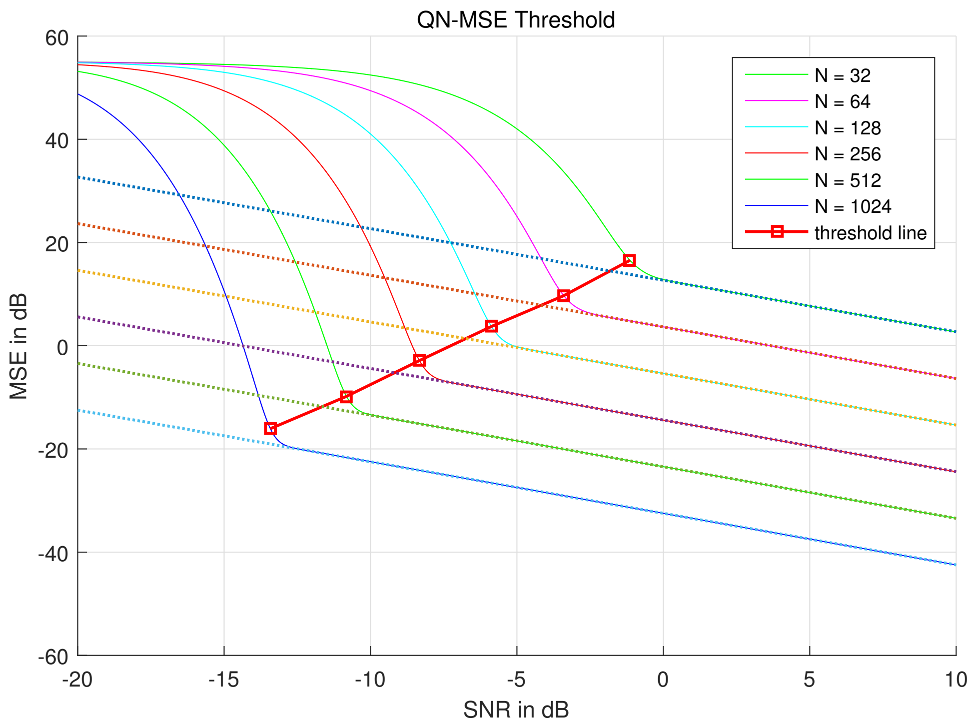

Figure 15: Q_N-MSE Threshold Expression

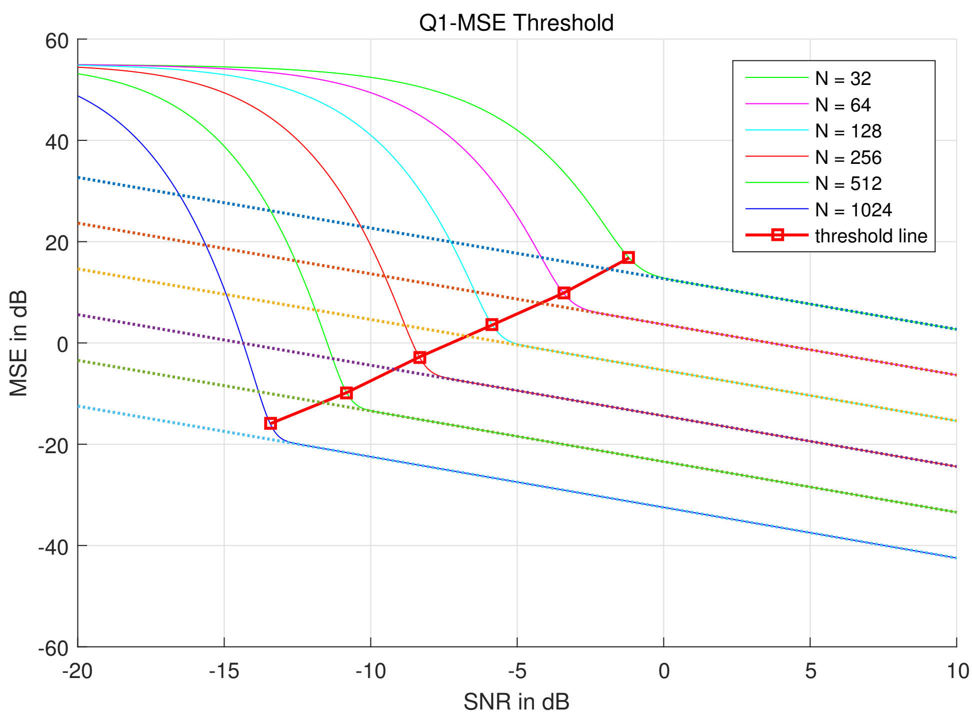

Figure 16: Q_1-MSE Threshold Expression

8.1 Accuracy of the New Expressions

As can be seen from the plots, thresholds given by the Q_N-MSE and Q_1-MSE are very close to each other. The expressions are compared with the actual 3dB turning points and Q&A threshold. At a N-value of 1024, the Q_N-threshold is at -13.4292 dB, the Q_1-threshold is at -13.4349 dB, the Q&A threshold is at -13.4331 and the actual turning point is at -13.433 dB. When N has increased to as large as 4096, the Q_N-threshold is at -18.1779 dB, the Q_1-threshold is at -18.1897 dB, the Q&A threshold is at -18.7091 and the real turning point is at -18.680, which have a difference less than 1dB. We continue searching for the point where the new threshold starts to suffer significantly from asymptotic deterioration. At an N-value of 5820 the difference between Q&A and Q_N-threshold exceeds 1.0 dB, marking the beginning of deteriorated accuracy. The normal choices of N are less than 1000 for the sake of lower estimate latency. Thus, for normal circumstances this expression is well capable of providing reliable threshold references.

We proceed to compare the root mean square difference between the actual 3dB deviation point and the results given by each expression. For integer N-values from 16 to 1024, the RMSE calculated to be 0.0126 (0.0283) dB, 0.0192 and 0.0157 dB for Q_N-MSE curve-fitting expression, Q_1-MSE curve-fitting expression and Q&A expression respectively. All expressions are stable with a very low root mean square error. Q_N-MSE curve-fitting expression is slightly more stable than the Q&A threshold while Q_1-MSE curve-fitting expression involves slightly more fluctuation. Thus, it can be concluded that the expression from curve-fitting with Q_N-MSE can provide improves slightly in accuracy for small Ns and that all three expressions can provide very accurate locations of thresholds for normal practical usage.

8.2 Viability in Different Estimators

As has been discussed in Section 4.3, MLE is usually implemented by FFT-based estimators which comprises two stages, the coarse search stage and fine search stage. The first stage is technically the Maximum Bin Search behind which the algorithm is N-point FFT plus a search for maximum frequency bin. The threshold onset, as in Section 5, is actually the beginning of outlier occurrence. Now that outliers are defined to be the maximizer in MBS failing to search for true frequency, the threshold effect is mainly dictated by the number of samples in MBS rather than the fine search algorithms.

In the coarse search stage, because of the small value of N, the estimators give estimates that are very inaccurate and, if without fine search stage, cannot improve in estimation error as the SNR increases (Figure 17). The fine search continues to search for the true peak of the main lobe, of which the shape gets closer to an ideal sinc function and improve the estimation accuracy (Figure 18). That is to say, even though fine search algorithms vary from one another, threshold region of the estimators does not change if N in MBS remains unchanged. Knowing that the number of samples in the MBS stage is usually small, we can confidently find the accurate threshold with the new expression regardless of what type of unbiased fine search algorithms are implemented in the fine search stage. This evidenced by figure 17 and 18 in that the location of turning points with merely MBS and with MBS plus fine search are approximately the same.

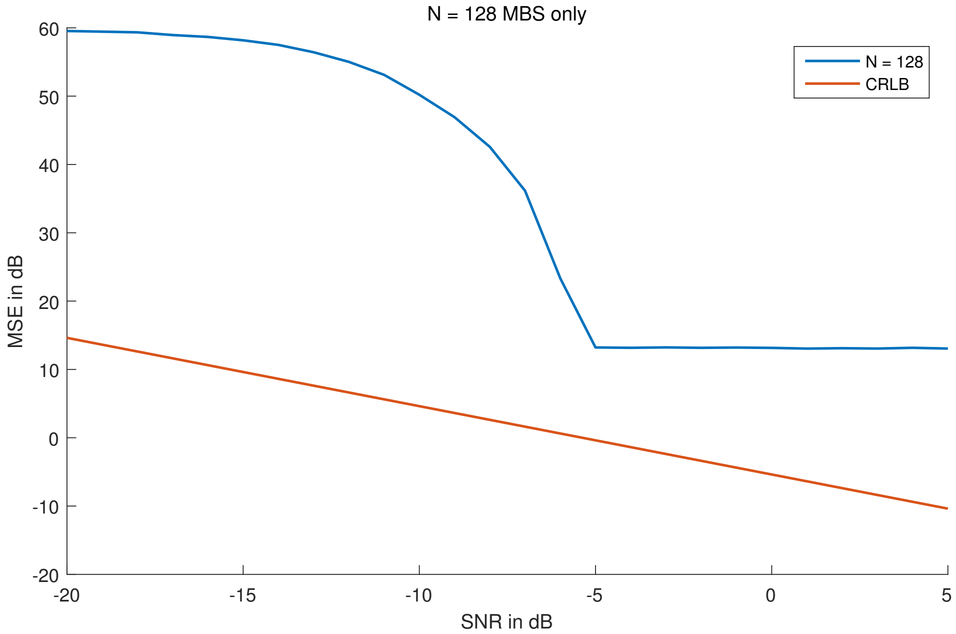

Figure 17: Maximum Bin Search (N = 128)

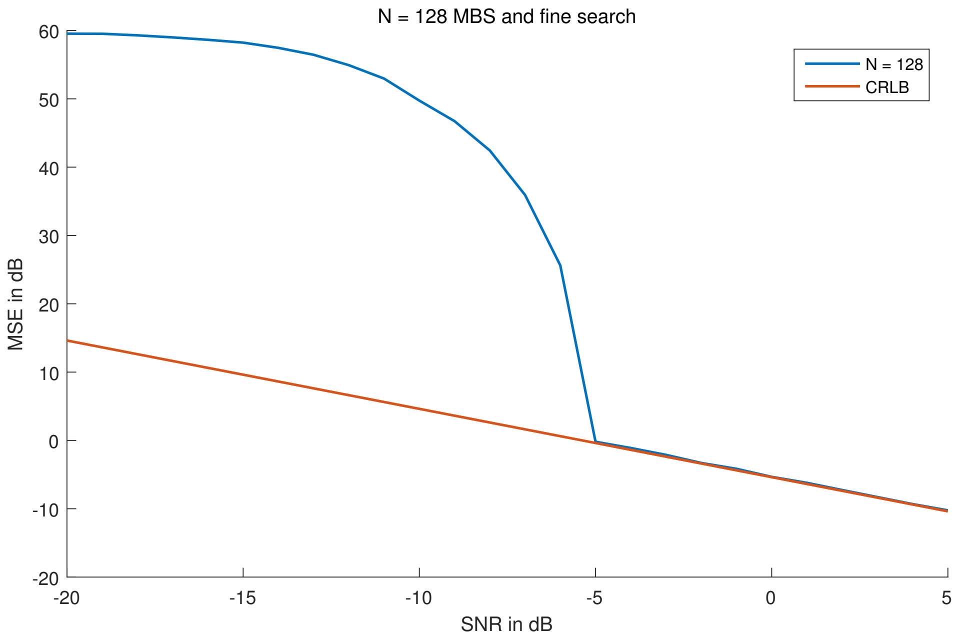

Figure 18: Maximum Bin Search and fine search (N = 128)

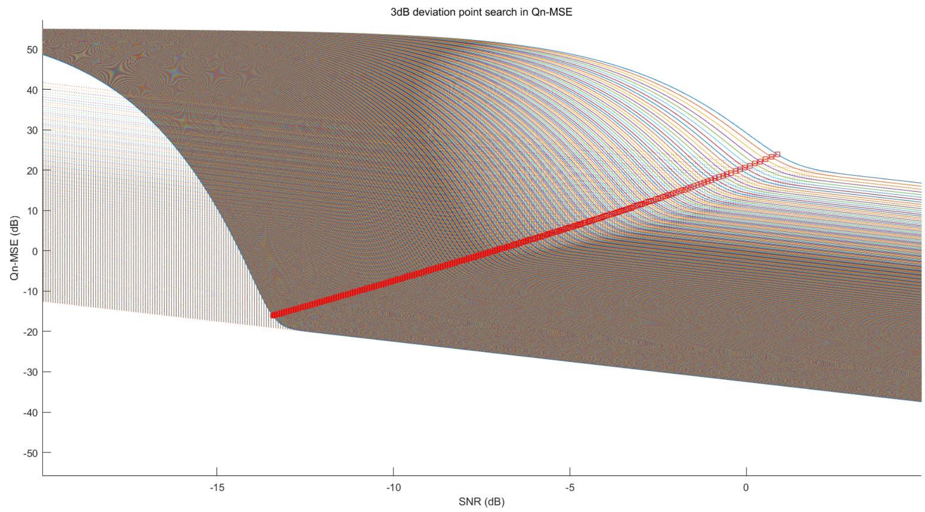

Figure 19: 3dB deviation points search (Q_N-MSE)

9 Conclusion

In this paper, we have outlined the phenomenon of threshold effect in unbiased MLE with the single tone and AWGN model. We have been researching theories concerning MLE and threshold effect and have formed a well-defined background basis.

Three different threshold expressions have been analyzed and compared with each other. Williamson et al's threshold utilizes the parabolic approximation of main lobe of Fourier transformed signal to set up a implementable estimator that can provide an accurate threshold indicator. The result is shown to be accurate with large number of samples, but when N is below 512 the indicator cannot accurately locate the 'knees' of MSE. Knockaert put forward a method to model the thresholds by the ratio of the tighter Barakin Bound against CRB. However, the classic Barakin Bound may not be tight enough to accurately locate threshold onset of the actual MSE and differs shape. The result is not very useful but the interpretation of threshold effect is highly valuable and it calls for the invention of more informative lower bounds in future research. Finally, the Q&A threshold is proposed from Rife and Boorstyn's model of outliers. It conquered the difficulty in handling non-closed form integrations in outlier probability expressions using convergent partitioning techniques. The result given by Q&A is shown to be very accurate and is chosen as the benchmark for the rest of research in this thesis.

Based on the benchmark of Q&A threshold, we have probed into the mechanism of threshold behavior from the angle of total and single outlier probabilities. After formulating the statistical model for the probability of exact numbers of outliers, we observed the overwhelming capability of single outlier in representing total outliers both theoretically and experimentally. Finally, a new threshold expression has been constructed using rational-polynomial approximation and has been compared with Q&A threshold. The expression by Q_N-MSE is shown to be more stable than Q&A threshold with N less than 1024 while that of Q_1-MSE suffers from slightly larger root mean sqare deviation from the actual 3dB turning points of MSE.

9.1 Suggestions for Future Research

Due to limitation of time and resources, my research has not reached conclusions in some other interesting directions. In future research, we could continue innovations of better lower bounds that may give previous information about the threshold effect. We could also relax some assumptions in Q&A threshold to make it more accurate and compatible with more signal models. It is also inspiring to use other possible techniques, such as Saddle Point Theory and Laplace Asymptotic Method, to degrade the outlier probability expressions to closed-form formulae.

References

[1] C. Johnson, W. Sethares, and A. Klein, *Software Receiver Design: Build Your Own Digital Communications System in Five Easy Steps*, 2011.

[2] W. Knight, R. Pridham, and S. Kay, “Digital signal processing for sonar,” *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1451–1506, 1981.

[3] B. Widrow and S. Stearns, “Adaptive signal processing,” Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, p. 491 , vol. 1, 1985.

[4] B. G. Quinn and P. J. Kootsookos, "Threshold behavior of the maximum likelihood estimator of frequency," *IEEE Transactions on Signal Processing*, vol. 42, pp. 3291-3294, 1994.

[5] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp.591-598, 1974.

[6] E. Aboutanios, “Iterative frequency estimation by interpolation on Fourier coefficients”, *IEEE Transactions on Signal Processing*, vol. 53, pp.1237-1241, 2005.

[7] L. Knockaert, "The Barankin bound and threshold behavior in frequency estimation," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2398-2401, 1997.

[8] R. C. Williamson, B. James and B. D. O. Anderson "Characterization of threshold for single tone maximum likelihood frequency estimation," *IEEE Transactions on Signal Processing*, vol. 43, pp. 817-821, 1995.

[9] Y. Liao, *Phase and Frequency Estimation: High-Accuracy and Low-Complexity Techniques.* PHD thesis, Worcester Polytechnic Institute, 2011

[10] J. Jacobson, *Lower Bounds on Estimator Error and the Threshold Effect.* thesis, Harvard College, 2004

[11] Pacific Northwest National Laboratory, *The Smart Grid: An Estimation of the Energy and CO2 Benefits,* PNNL, 2010.

[12] S. S. Haykin, *Communication systems,* 3rd ed. New York: Wiley, 1994.

[13] L.C. Lucien, *Maximum Likelihood – an Introduction,* ISI Review, 1990.

[14] Y. V. Zakharov and T. C. Tozer,*Frequency estimator with dichotomous search of periodogram peak,* Electron. Lett., vol. 35, pp. 1608–1609, 1999.

[15] E. Aboutanios, *Frequency Estimation for Low Earth Orbit Satellites*, 2003

[16] B.L. Kristine, S. Yossef, E. Yariv and L.V.T. Harry, "Extended Ziv-Zakai Lower Bound for Vector Parameter Estimation", *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 624-637, 1997.

[17] C. Ren, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Hybrid Lower Bound On The MSE Based On The Barankin And Weiss-Weinstein Bounds", ICASSP'2013: 38th International Conference on Acoustics, Speech, and Signal Processing, Vancouver, Canada, 2013.

A Required Number of Repeated Experiments

For an event of probability P, we would like to determine how many repeated experiments is sufficient for a 'smooth' observation of the event's occurrence. Considering solely the event itself, we can model it as a Bernoulli variable. The variance of estimated probability has the form of

$$\begin{aligned} Var(\hat{P}) &= \frac{NP(1-P)}{N^2} \\ &= \frac{P(1-P)}{N}. \end{aligned} \tag{63}$$

the standard variation of estimated probability can be given by

$$Std(\hat{P}) = \sqrt{\frac{P(1-P)}{N}}. \tag{64}$$

When we want the ratio (ϵ) of standard deviation of experiments to be small, such as 0.1,

$$\epsilon = \frac{Std(\hat{P})}{P} = 0.1. \tag{65}$$

Solving for N we have

$$\begin{aligned} N &= \frac{1-P}{P} \cdot \frac{1}{\epsilon^2} \\ &= 100\frac{1-P}{P} \end{aligned} \tag{66}$$

For a small P this approximates to

$$N = \frac{100}{P} \tag{67}$$

In light of this theory, when our outlier probability is as small as 10^{-5}, the number of experiments would have to surpass 10^7, which is 10 million. According to my experimental results, for small probabilities, an ϵ of 0.1 is still not satisfactorily small so that 100 million repeated experiments would be sufficient to beautifully observe this event.