Undergraduate Project

Order Estimation of Parameters of Signal Processing Models

Supervisor : Professor Amit Mitra
Co-supervisor: Professor Ketan Rajawat
Student Name: Chirumamilla Satya Keerthana, 210290

Acknowledgement

I want to convey my heartfelt appreciation and gratitude to Professor Amit Mitra for providing me with the chance to work on this project and for his wonderful guidance throughout the project.

I would also like to thank Professor Ketan Rajawat for acting as my co-supervisor for the project and giving me a chance to take up UGP

Contents

Introduction

Various Methods of Order Estimation

Results

Future Work : Robust Methods

References

Introduction

Need of the proposed problem statement :

Order estimation in signal processing is necessary to describe the signal accurately to capture the complexity of the signal without losing much information and without increasing the complexity of the model too much.

The order in the problem statement refers to the approximation to number of source signals in the original signal.
We will restrict ourselves to the sinusoidal and chirp signal models

A good model is that which is resistant to noise in the signal and avoids overfitting for the outliers in the data.

Various Methods in the Order Estimation

1) MAP
2) BIC
3) Corrected BIC
4) AIC
5) Corrected AIC
6) PAL

Various Methods of Order Estimation

Notations for used :

y = the vector of available data (of size N)
θ = the (real-valued) parameter vector
n = the dimension of θ.

$$\hat{\theta} = \arg\max_{\theta} \ln p(y, \theta)$$

$p(y, \theta) =$ the probability density function (PDF)

H_n denote the hypothesis that the model order is n

$p_n(y|H_n) =$ the PDF of y under H_n.

$p_n(H_n) =$ a priori probability of H_n

$p_n(H_n|y) =$ conditional probability of H_n

let $\overline{n}$ denote a known upper bound on n, $n \in [1, \overline{n}]$

MAP (Maximum A Posteriori) :

It is a bayesian approach which includes the prior information along with the likelihood function. It aims to find order that maximizes the posterior probability with the given data.

MAP simplifies to maximizing the given probabilities:

$$p_n(H_n|y) = \frac{p_n(y|H_n)p_n(H_n)}{p(y)}. \quad \max_{n \in [1, \overline{n}]} p_n(y|H_n)p_n(H_n).$$

AIC (Akaike Information Criterion) :

AIC's order selection is based on the rule which incorporates both likelihood function and penalizing the model if the model tries to overfit.

$$\text{AIC} = -2 \ln p_n\left(y, \hat{\theta}^n\right) + 2n$$

Picture credits: [1]

Various Methods of Order Estimation

Corrected AIC:

Corrected AIC has a higher penalty term compared to AIC and hence is better compared to AIC as it doesn't try to overfit unlike AIC

$$\mathrm{AIC_c} = -2 \ln p_n(y, \hat{\theta}^n) + \frac{2N}{N-n-1} n$$

As $N \to \infty$, $\mathrm{AIC_c} \to \mathrm{AIC}$

BIC (Bayesian Information Criterion) :

In BIC, the penality term grows logarithmically with number of observations and hence avoids complex models

$$\mathrm{BIC} = -2 \ln p_n(y, \hat{\theta}^n) + n \ln N$$

Corrected BIC :

The objective function has slight modification compared to BIC where it accounts for small sample sizes. It is more suitable when the sample size is small.

GIC (General Information Criterion) :

It refers to the family of model order selection. Most common used are AIC, BIC, HQIC. All of these differ in the penalty terms.

$$\mathrm{GIC} = -2 \ln p_n(y, \hat{\theta}^n) + \nu n$$

values of ν in the interval $\nu \in [2, 6]$ have been found to give the best performance

Picture credits: [1]

Various Methods of Order Estimation

PAL (Penalty Adjusted Likelihood) :

PAL method penalizes adaptively based on the data. The penalty is small for order selection less than true order and higher for orders greater than the true order.

Notations used for this definition:

Let $\boldsymbol{y} \in \mathbb{R}^N$ denote the data vector.

$p_n(\boldsymbol{y}, \boldsymbol{\theta}_n)$ is called the likelihood function

$\boldsymbol{\theta}_n$ is the vector of unknown parameters

$\hat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimate of $\boldsymbol{\theta}_n$

PAL method maximizes the following term:

$$\mathrm{PAL}(n) = -2 \ln p_n(\boldsymbol{y}, \hat{\boldsymbol{\theta}}_n) + n \ln(\tilde{n}) \frac{\ln(r_n + 1)}{\ln(\rho_n + 1)}$$

where

$$r_n = 2 \ln \left[\frac{p_{n-1}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}_{n-1})}{p_1(\boldsymbol{y}, \hat{\boldsymbol{\theta}}_1)} \right]$$

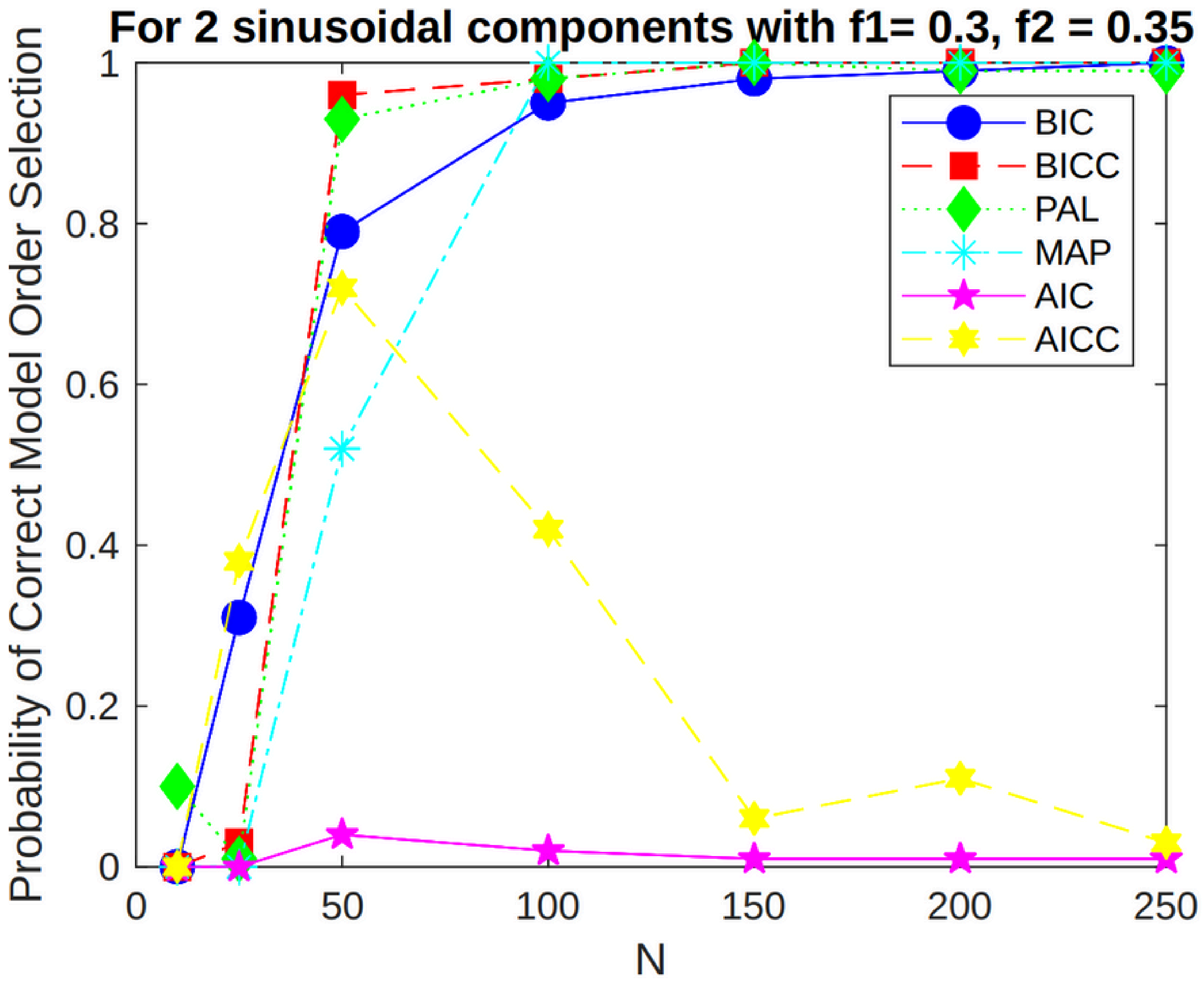
$$\rho_n = 2 \ln \left[\frac{p_{\tilde{n}}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}_{\tilde{n}})}{p_{n-1}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}_{n-1})} \right]$$

Picture credits: [3]

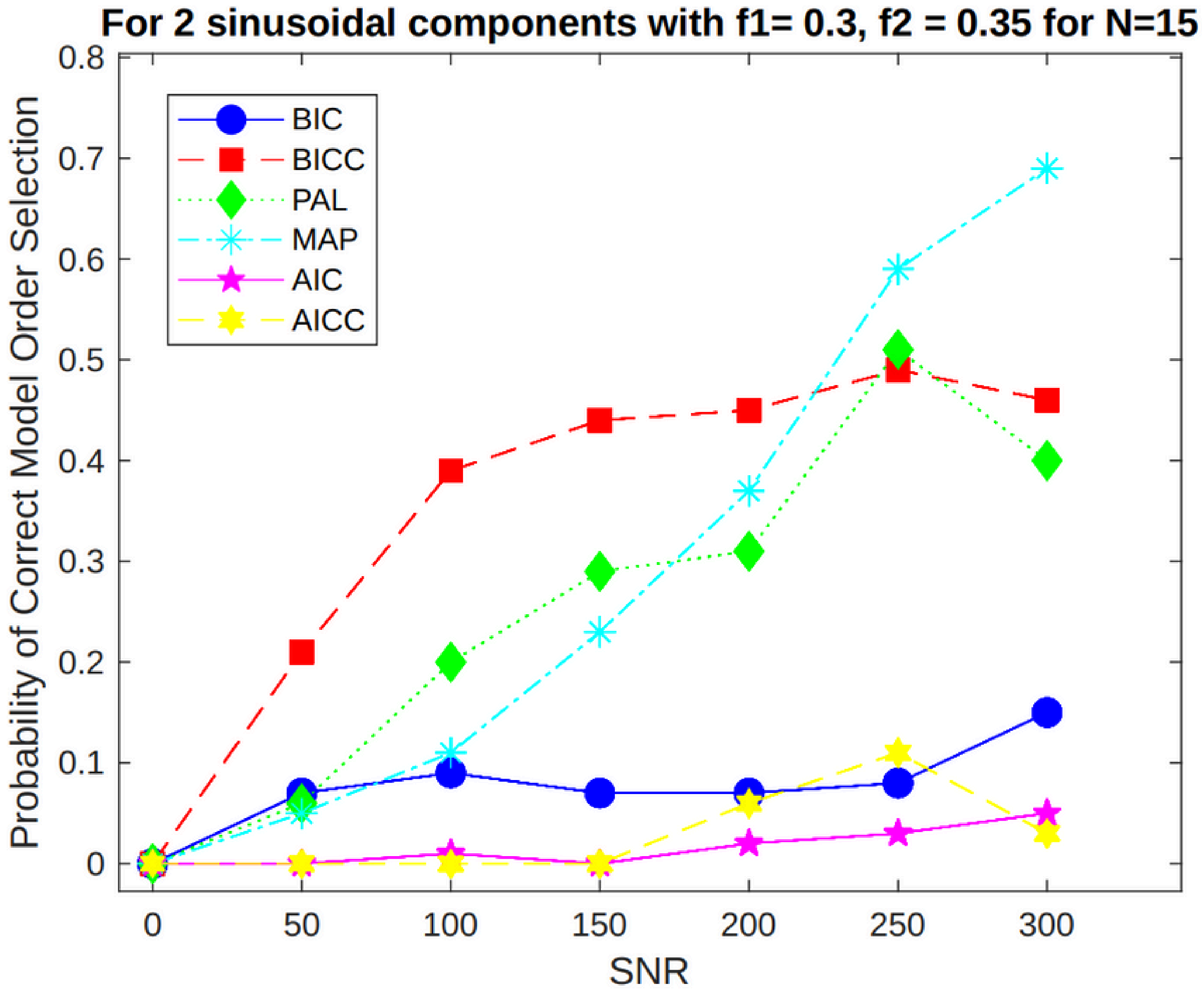
Results :

A. Noise following Normal distribution:

Probability Vs N:



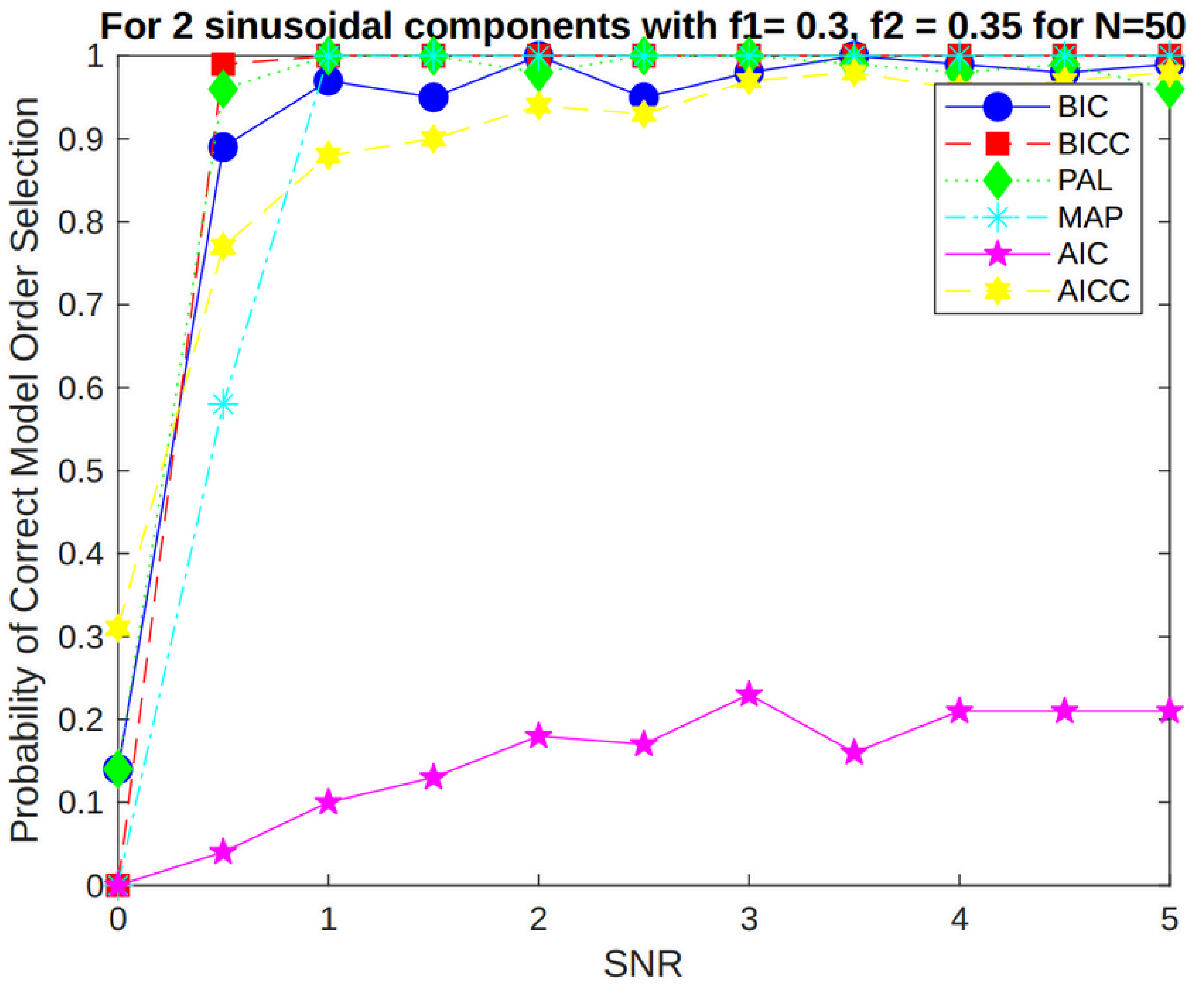
Probability Vs SNR for N=15



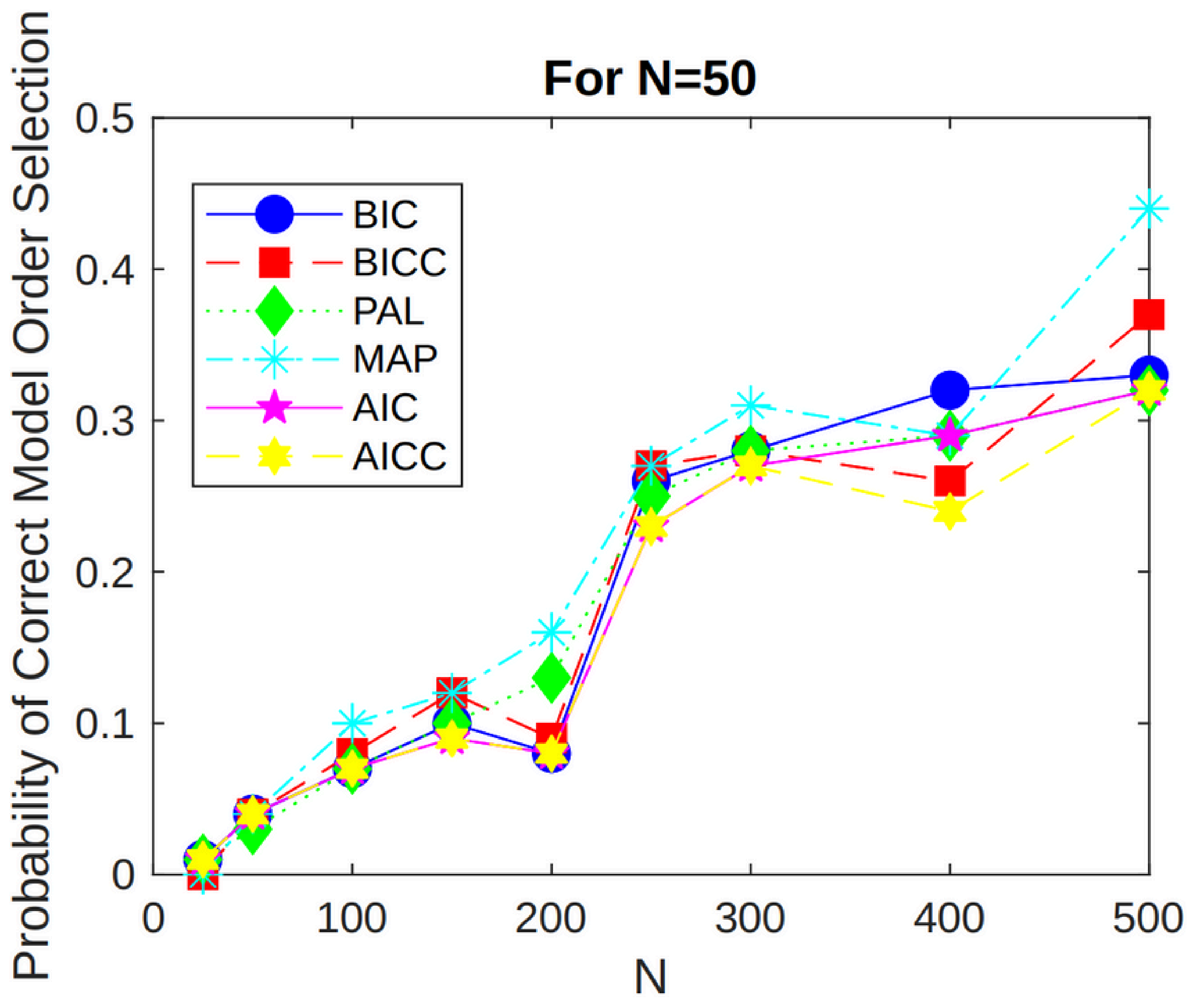
For all the graphs, number of simulations is 100 and N is the sample size.
SNR is inversely proportional to square of standard deviation.

Results :

Probability Vs SNR for N=50

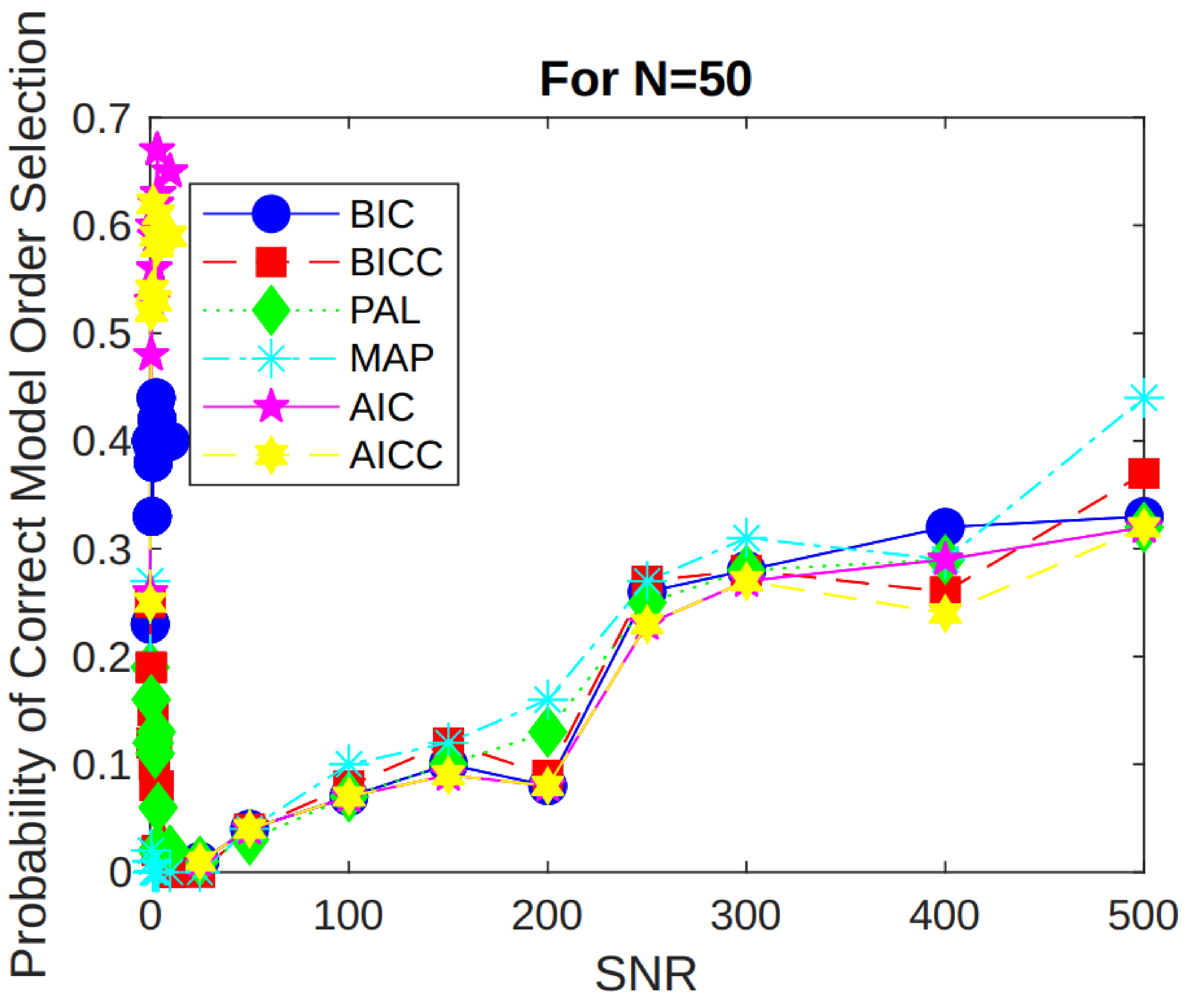


Probability Vs N for chirp signal model :



Results :

Probability Vs SNR for a chirp signal model



Results :

No. of times correct order was selected in 100 simulations Vs Difference in frequencies

For Sinusoidal Model:

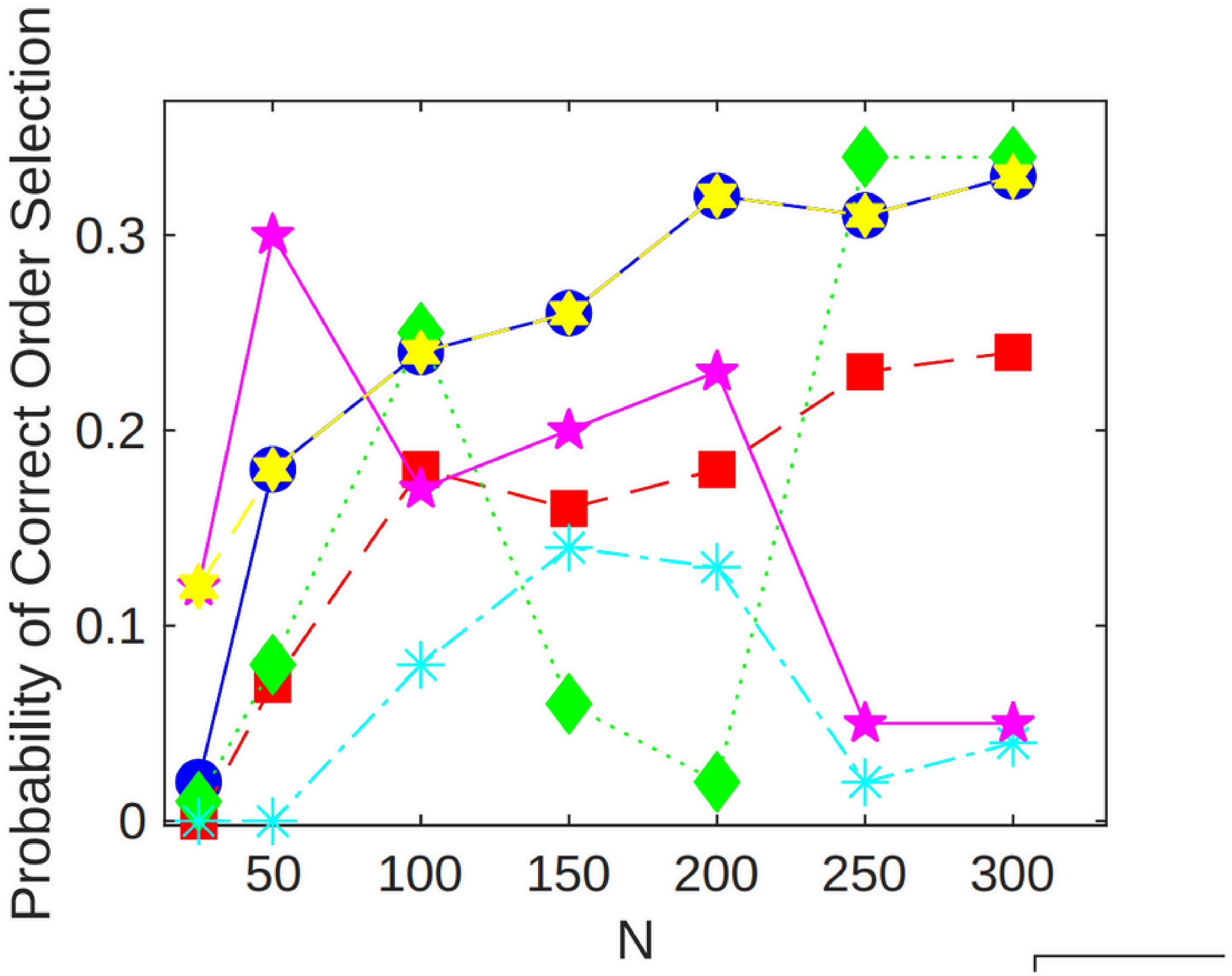
Model/ Difference In frequencies	10	5	1	0.5	0.1	0.05	0.01	0.001	0.0001
BIC	65	67	85	96	100	100	20	18	18
BICC	98	99	100	100	100	100	1	0	0
PAL	98	99	100	100	100	94	0	0	0
MAP	100	100	100	100	100	100	0	0	0
AIC	0	1	3	6	85	27	0	0	1
AICC	40	33	66	88	99	99	28	33	28

With these simulations, we show how the models perform when the frequencies are close to each other.

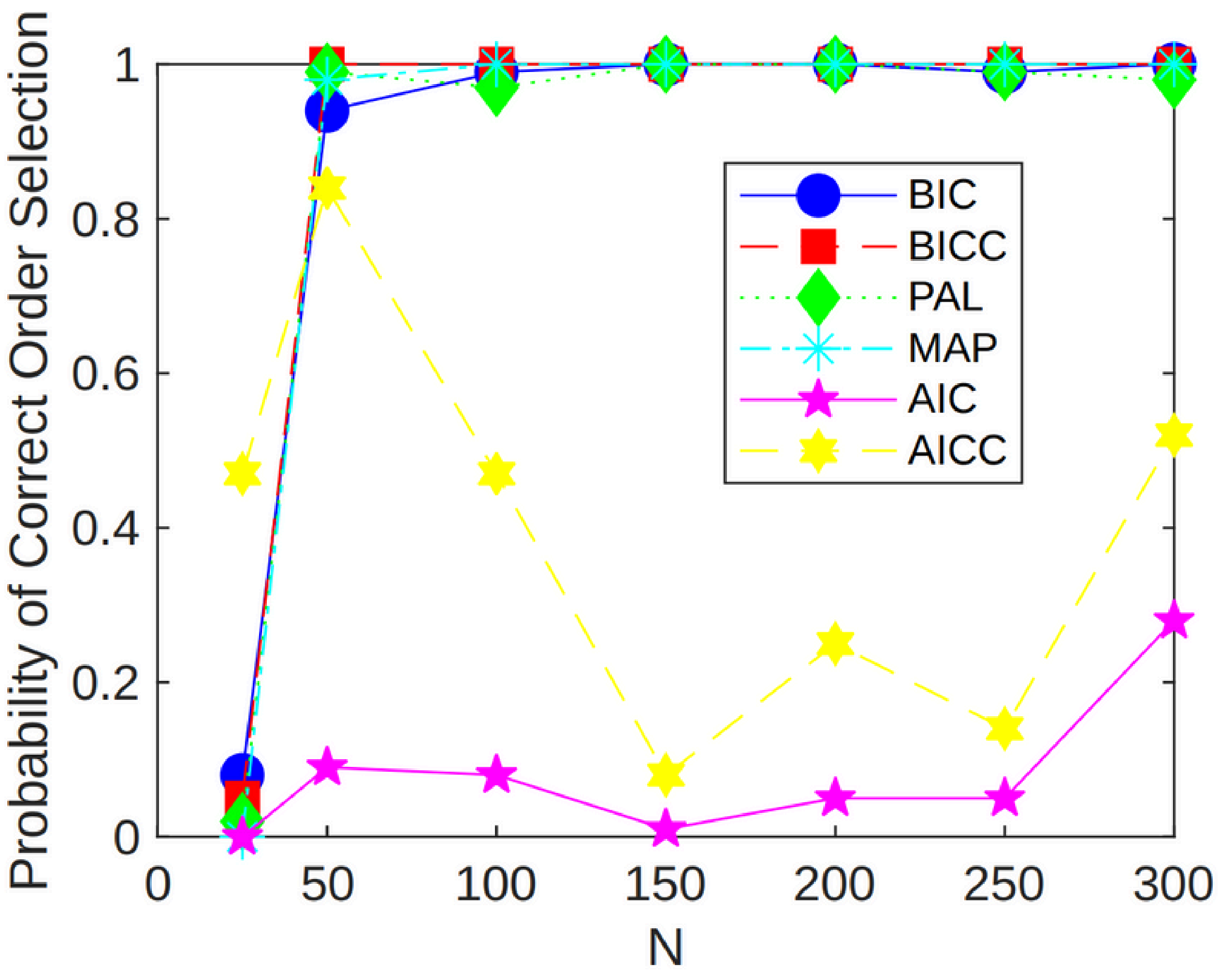
Results :

B. Noise following t-distribution: (sinusoidal signal model)

With degrees of freedom = 1
(Cauchy distribution, heavy tailed)



With degrees of freedom = 15



The methods are non-robust for heavy tailed noise. With higher degrees of freedom, t-distribution has a small tail

Future Work : Robust Methods

We find that the all the models we have considered so far are non-robust, i.e, they are sensitive to outliers and any deviation from the assumptions gives poor results. Hence, methods which are less sensitive to outliers are needed. Some such methods are based on the stratified bootstrap approach.

Bootstrap Technique :

Bootstrap involves resampling the data to estimate the distribution of sample. It makes no assumptions about type of error, hence is less sensitive to outliers.

It involves calculation of confidence intervals which are not dependent on assumptions made for special distributions. This is a robust method to handle data with noise such as t-distribution (heavy tailed)

This is useful when the sample size is small as it creates large data by resampling, hence robust.

Some Applications:

1) Choosing the error function in Linear Regression for robustness [5]

2) Stratified Bootstrap approaches for robust model order estimation [6]

References :

[1] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," in IEEE Signal Processing Magazine, vol. 21, no. 4, pp. 36-47, July 2004

[2] Müller, Samuel, and Alan H. Welsh. “On Model Selection Curves.” International Statistical Review / Revue Internationale de Statistique, vol. 78, no. 2, 2010

[3] Petre Stoica, Prabhu Babu, Model order estimation via penalizing adaptively the likelihood (PAL), Signal Processing, Volume 93, Issue 11, 2013, Pages 2865-2871, ISSN 0165-1684

[4] Khushboo Surana, Sharmishtha Mitra, Amit Mitra, Petre Stoica, Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties, Signal Processing

[5] Outlier Robust Model Selection in Linear Regression, Samuel MÜLLER and A. H. WELSH

[6] Model Selection: A Bootstrap Approach, A. M. Zoubir, Communications and Information Processing Group School of Electrical & Electronic Systems Engineering Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia

Thank you!