# Model order estimation in multi-component chirp signals

Abhinek Shukla[2], Debasis Kundu[1], and Amit Mitra[1]

[2]Department of Statistics & Data Science, National University of Singapore, Singapore 117546
[1]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur - 208016, India
[2]Corresponding author. Email: abhushukla@gmail.com

**Abstract**

Real datasets exhibiting non-constant frequencies, can be analysed with the help of multi-component chirp signals adequately. This requires estimation of the number of chirp components in practice. The problem of estimation becomes difficult when the data is composed of moderate to large number of chirp components. We propose Bayesian approach to estimate the number of components of a multi-component chirp model using an approximation to the mode of the posterior distribution. This approximation is motivated from the relation between mode of the obtained posterior and the sequential least squares estimators. Extensive numerical simulations highlight superior performance of the proposed estimators than that of the estimators based on Akaike information criterion, Bayesian information criterion and asymptotic maximum a posteriori probability, for a large range of signal-to-noise ratio levels. We further illustrate the implementation of the proposed estimators through two real data examples.

**Keywords:** Posterior distribution, Gaussian errors, multi-component chirp, non-linear regression, order estimation.

## 1 Introduction

Multi-component chirp signals can be used to model datasets arising in many engineering applications, such as synthetic aperture radar (SAR) imaging of speeding uncooperative targets with multi-scatterers and multipath propagation in mobile telecommunications. Depending on these applications, the number of signal components can represent the number of dominant scatterers in SAR case or the number of multiple paths in mobile communications case.However, the number of components is usually unknown for real life data, see [4] and [15]. In present article, we have considered the estimation of the number of components for 1-D multi-component chirp models. For the considered problem, the interpretation for model order estimation is to find minimum set of dominant frequency and frequency rates, which explains "most" of the variations or patterns in the received signal. The problem of estimation of the number of components is a particular case of the problem of model selection from a set of candidate nested models and the existing literature is vast, e.g., see [1, 2, 7, 19] and [20].

While selecting an appropriate or good model order, we know that increasing the number of components in the model to fit a given data will improve the fitting, however we will have to pay with complexity of the model due to large number of components parameters. Generally, as the number of components will grow, bias will decrease however the variance will increase, see [1]. The model with larger number of components will be less flexible and is expected to work poorly for new future data points. So, one of the important aim would be to somehow balance this increased performance of fitting versus the increased complexity. Maybe a more accurate measure of a good model lies in its ability to perform well on the current data from which the model is estimated and also on the future data sets obtained from the same process. A model that is overly complex and fits the current data set excellently might not fit subsequent data sets satisfactorily. Conversely, a model that is overly simplistic may fail to fit any of the data sets effectively.

Using classical methods of model selection, we get a single estimate of the number of components for a given data. We have considered a specific problem of finding the set of "most" dominant frequency and frequency rates pair, which can explain the data satisfactorily. In our problem, proposing a method which can present a range of choices for the model order along with being flexible enough, will gain more response from the practitioners than drawing inference based on a single value of estimate of the model order. Also, sometimes an engineer or practitioner may have a good guess for the model order from the beginning, before even observing the data, in those situations, incorporating this knowledge itself will result in capturing data patterns better.

Therefore, it is natural to estimate the number of components using Bayesian methods, as the posterior distribution helps quantifying uncertainty around the number of components. When the number of components of the data generating model is possibly moderate to large, the Bayesian methods will require computation of high-dimensional integrals, however, the posterior of the unknown parameters may not admit any closed-form analytical expression when the frequency or frequency rates are unknown. Further, when the non-linear parameters of different components are well-separated or the sample size available is sufficiently large, then the two major strategies are: (i) analytic approximations to the integrals; (ii) posterior expansion around the maximum likelihood estimate of the non-linear parameters. The validity of these approximations are difficult to assess in the interesting cases where the amount of available data is small or non-linear parameters of different components are close, see e.g., [3].

Many articles can be found on Bayesian parameter estimation of multi-component chirp signal models with known number of components, e.g., [9], [18], [22] and [27]. However, for unknown number of components, the literature is relatively less, e.g., [8], [15] and [26]. Estimators of [15] were based on high order ambiguity function which is asymptotically sub-optimal for the frequency and frequency rate parameters, further the error accumulated in estimation of higher order parameters based on these techniques, is propagated and affects seriously the estimation error of lower order parameters. These effects directly lead to the inferior performance of the estimation of the number of components.

Methods based on reversible jump Markov chain Monte Carlo (RJMCMC) have been proposed in

[8], however, these algorithms suffer from slow convergence due to low acceptance probabilities for across model moves, e.g., see [5] and [12]. The degree of confidence on convergence also deteriorates very rapidly as the dimension of the state space increases. The algorithm based on RJMCMC is difficult to implement in a parallel fashion because the algorithm produces only one sample at each step, where this sample depends on the sample produced in the former step [14].

An adaptive importance sampling (AIS) based sampling procedure, population Monte Carlo (PMC) is utilized in [26] for estimating the number of chirp components. These methods were based on detecting certain peaks of the modified discrete chirp Fourier transform, which, by definition, provides only limited parameter space $(0, 1/N)$ for the chirp rate parameter if we compare with the coefficient of the $n^2$ term in (1) with that of considered in [26]. PMC performs sampling from the exact posterior distribution providing unrestricted moves as compared to the RJMCMC, however this leads to increased storage requirements. Further, the performance of the AIS method gravely depends on the selection of the importance function. The resampling step of the methods belonging to the PMC family also causes the path degeneracy see e.g., [11] and hence, jeopardizing the diversity of the proposals in each of the iterations, which results in less exploration of the state space.

Model order estimation can also be performed based on Akaike information criteria (AIC) or Bayesian information criteria (BIC). However, it is well known that the conventional AIC tends to overestimate the model order due to weak penalization, e.g. see [24], and it mostly chooses the most available complicated model. It has also been observed in the literature that AIC gives inconsistent estimates of the model order [16]. BIC on the other hand, usually tends to underestimate and performs poorly for low signal-to-noise ratio (SNR) levels.

In this article, we propose a method to estimate the number of chirp components based on approximating the mode of the joint posterior distribution of the number of components and the non-linear parameters, by utilizing the sequential least squares estimator (LSE). Our proposed estimator neither requires sampling from the posterior distribution nor expanding the posterior around the mode or maximum likelihood estimate. Extensive numerical simulations suggest that the sequential LSE of the non-linear parameters can help approximating the exact posterior of the number of components for a very large range of SNR.

The paper is organized as follows. We introduce the model, the choice of suitable priors, derive the posterior distribution, and then the proposed estimation methodology in Section 3. Numerical performance of the proposed estimator and its comparison with AIC, BIC and the asymptotic maximum a posteriori (AMAP) rule [10] based estimators is shown in Section 4. Further, two real datasets are analyzed with the proposed estimators and compared with standard methods in Section 5, followed by the conclusion in Section 6 and appendix in Section 7.

## 2 Model Introduction

We consider the following general real-valued multi-component chirp signal model as:

$$y(n)=\sum_{\ell=1}^{K}\Big(A_\ell\cos(\alpha_\ell n+\beta_\ell n^2)+B_\ell\sin(\alpha_\ell n+\beta_\ell n^2)\Big)+X(n),\ \ n=1,2,\ldots,N, \tag{1}$$

where, $X(n)$ is an i.i.d. sequence of Gaussian errors with mean 0 and variance $\sigma^2$, $K$ is the number of components, $A_\ell$ and $B_\ell$ are the amplitude parameters, $\alpha_\ell$ is the frequency and $\beta_\ell$ is the frequency rate parameter of the $\ell^{th}$ component, $\ell=1,2,\ldots,K$ and $N$ is the sample size. Each parameter is assumed to be unknown in the model (1). The non-linear parameters $\alpha_\ell$ and $\beta_\ell$ can be used to analyze the speed and acceleration of the moving object respectively, see [4]. The size and smoothness of the surface of the target can be studied as a function of the amplitude parameters, see e.g., [26].

### 2.1 Definition of AIC and BIC criterion

We consider the standard methods of model order estimation Akaike information criterion (AIC) and Bayesian information criterion (BIC) which belong to information criterion family and work on using data to assign each candidate model a score and then find the best score candidate model as the selected model, see [7]. AIC for a model is generally defined (see [20] and [1] )as follows:

$$\text{AIC}=-2\times\log\left(\text{likelihood}\right)+2\times\text{number of parameters}.$$

Mathematically, if we denote $\log(L(\widehat{\boldsymbol{\theta}}))$ as the log-likelihood evaluated at the maximum likelihood estimators, then AIC takes the following form for the considered model (4) with $\boldsymbol{\theta}$ unknown parameter and $k$ number of components:

$$\text{AIC}(k)=-2\times\log(L(\widehat{\boldsymbol{\theta}}))+2\times 4k.$$

Here $4k$ appears because for each component we have 4 parameters. In our case for the i.i.d. normal errors with mean 0 and variance $\sigma^2$ in the model (4), the log-likelihood is given as:

$$\log(L(\boldsymbol{\theta}))=\frac{-N\log(2\pi)}{2}-\frac{N\log(\sigma^2)}{2}-\frac{1}{\sigma^2}\Big(\boldsymbol{Y}-\boldsymbol{D}(\boldsymbol{\eta}_k)\boldsymbol{\mathcal{A}}_k\Big)^\top\Big(\boldsymbol{Y}-\boldsymbol{D}(\boldsymbol{\eta}_k)\boldsymbol{\mathcal{A}}_k\Big).$$

Here $\boldsymbol{\theta}=(\boldsymbol{\mathcal{A}}_k,\boldsymbol{\eta}_k,\sigma^2)^\top$, now for the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}$, we have

$$\log(L(\widehat{\boldsymbol{\theta}}))=\frac{-N\log(2\pi)}{2}-\frac{N\log(\widehat{\sigma}^2)}{2}-N,$$

because MLE of $\sigma^2$ is $\widehat{\sigma}^2=\dfrac{\Big(\boldsymbol{Y}-\boldsymbol{D}(\widehat{\boldsymbol{\eta}}_k)\widehat{\boldsymbol{\mathcal{A}}}_k\Big)^\top\Big(\boldsymbol{Y}-\boldsymbol{D}(\widehat{\boldsymbol{\eta}}_k)\widehat{\boldsymbol{\mathcal{A}}}_k\Big)}{N}$.

Then, $\text{AIC}(k)$ takes the form

$$\text{AIC}(k)=N\log(2\pi)+N\log(\widehat{\sigma}^2)+2(4k),$$

or equivalently, after rearranging the constant terms free from $k$ and scaling by 2, we can write AIC criterion as follows:

$$\mathrm{AIC}(k) = N \log(N\widehat{\sigma}^2)/2 + 4k.$$

Note that obtaining LSE is computationally challenging, so we need some other estimator which has same asymptotic performance as that of the LSE. We use sequential LSE [17] instead of LSE to compute the residual sum of square. We denote residual sum of square by $\mathrm{RSS}(k)$ which is obtained by fitting a $k$ component chirp model based on sequential LSE $\widetilde{\boldsymbol{\eta}}_k$ and $\widetilde{\boldsymbol{\mathcal{A}}}_k$. So, the form of AIC considered in the thesis for comparison is provided in (14). Similarly, we can obtain form of BIC as in (15) based on its following general definition:

$$\mathrm{BIC}(k) = -2 \times \log\left(\text{likelihood}\right) + \log(N) \times \text{number of parameters}.$$

Next we illustrate the sequential estimation of multi-component chirp model parameters (1) to obtain $\mathrm{RSS}(k)$ in following steps:

- **Step-1** Consider the following sum of squares:

  $$Q_1(\boldsymbol{\theta_1}) = \sum_{n=1}^{N} \left(y(n) - A_1 \cos(\alpha_1 n + \beta_1 n^2) - B_1 \sin(\alpha_1 n + \beta_1 n^2)\right)^2.$$

  First component parameter vector is $\boldsymbol{\theta}_1^0 = (A_1^0, B_1^0, \alpha_1^0, \beta_1^0)^\top$ and let $\boldsymbol{\theta}_1 = (A_1, B_1, \alpha_1, \beta_1)^\top$, then we define sequential LSE of the first component parameters as follows:

  $$\widetilde{\boldsymbol{\theta}}_1 = \arg\min_{\boldsymbol{\theta}_1} Q_1(\boldsymbol{\theta}_1). \tag{2}$$

- **Step-2** We eliminate the effect of first component from the data and then define sequential LSE of second component as $\widetilde{\boldsymbol{\theta}}_2 = (\widetilde{A}_2, \widetilde{B}_2, \widetilde{\alpha}_2, \widetilde{\beta}_2)^\top$ of second component parameter vector $\boldsymbol{\theta}_2^0 = (A_2^0, B_2^0, \alpha_2^0, \beta_2^0)^\top$ by updating data as:

  $$\widetilde{y}_1(n) = y(n) - \widetilde{A}_1 \cos(\widetilde{\alpha}_1 n + \widetilde{\beta}_1 n^2) - \widetilde{B}_1 \sin(\widetilde{\alpha}_1 n + \widetilde{\beta}_1 n^2).$$

  Consider,

  $$Q_2(\boldsymbol{\theta}_2) = \sum_{n=1}^{N} \left(\widetilde{y}_1(n) - A_2 \cos(\alpha_2 n + \beta_2 n^2) - B_2 \sin(\alpha_2 n + \beta_2 n^2)\right)^2,$$

  where, $\boldsymbol{\theta}_2 = (A_2, B_2, \alpha_2, \beta_2)^\top$, then the second component estimator is given by the following equation:

  $$\widetilde{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}_2} Q_2(\boldsymbol{\theta}_2). \tag{3}$$

- **Step-3** We repeat Step-2 up to $k-2$ number of times to get the sequential LSE of parameters of $k$ component chirp model (1) with $K = k$.

- **Step-4** Once we have the fitted data for $k$ component chirp model as

$$\widetilde{y}_k(n) = \sum_{\ell=1}^{k} \Big(\widetilde{A}_\ell \cos(\widetilde{\alpha}_\ell n + \widetilde{\beta}_\ell n^2) + \widetilde{B}_\ell \sin(\widetilde{\alpha}_\ell n + \widetilde{\beta}_\ell n^2)\Big),$$

then we can obtain $\text{RSS}(k) = \sum_{n=1}^{N} \big(y(n) - \widetilde{y}_k(n)\big)^2$.

However, the classical methods AIC and BIC based estimators of model order do not perform well, specifically, AIC suffers from overestimation of the model order and BIC suffers from poor performance under low SNR levels, as will be seen in later part of simulations results. We propose our estimators in following subsections, based on Bayesian methodology by defining a set of priors on the model parameters and studying the posterior distribution.

## 3 Proposed Methodology

Let us reformulate equation (1) in the following vector and matrix form:

$$\boldsymbol{Y} = \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\boldsymbol{\mathcal{A}}_K + \boldsymbol{X}, \tag{4}$$

where $\boldsymbol{Y} = (y(1), y(2), \ldots, y(N))^\top$ is the data vector, $\boldsymbol{X} = (X(1), X(2), \ldots, X(N))^\top$ is the noise vector, $\boldsymbol{\eta}_K = (\boldsymbol{\alpha}_K^\top, \boldsymbol{\beta}_K^\top)^\top$ is the vector of non-linear parameters, $\boldsymbol{\alpha}_K = (\alpha_1, \alpha_2, \ldots, \alpha_K)^\top$ is the vector of frequency parameters and $\boldsymbol{\beta}_K = (\beta_1, \beta_2, \ldots, \beta_K)^\top$ is the vector of frequency rate parameters, $\boldsymbol{\mathcal{A}}_K = (A_1, B_1, A_2, B_2, \ldots, A_K, B_K)^\top$ is the amplitude vector. $\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)$ is a $N \times 2K$ order matrix with following expression:

$$\begin{bmatrix} \cos(\alpha_1+\beta_1) & \sin(\alpha_1+\beta_1) & \cos(\alpha_2+\beta_2) & \sin(\alpha_2+\beta_2) & \cdots & \cos(\alpha_k+\beta_k) & \sin(\alpha_k+\beta_k) \\ \cos(2\alpha_1+2^2\beta_1) & \sin(2\alpha_1+2^2\beta_1) & \cos(2\alpha_2+2^2\beta_2) & \sin(2\alpha_2+2^2\beta_2) & \cdots & \cos(2\alpha_k+2^2\beta_k) & \sin(2\alpha_k+2^2\beta_k) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(N\alpha_1+N^2\beta_1) & \sin(N\alpha_1+N^2\beta_1) & \cos(N\alpha_2+N^2\beta_2) & \sin(N\alpha_2+N^2\beta_2) & \cdots & \cos(N\alpha_k+N^2\beta_k) & \sin(N\alpha_k+N^2\beta_k) \end{bmatrix}.$$

We will be treating the estimation problem of the unknown parameter $K$ along with estimation of other parameters $(\boldsymbol{\mathcal{A}}_K, \boldsymbol{\alpha}_K, \boldsymbol{\beta}_K)$ as the by-product. We will interpret the estimate of the model order as the most probable number of components for the given data with the corresponding residual errors being i.i.d. Gaussian with mean 0.

### 3.1 Prior on the Number of Components

We consider diverse choices for the prior on $K$ as:
(i) truncated Poisson distribution with parameter $\lambda$, denoted as Pois($\lambda$) i.e.,

$$p_K(k) = \mathbb{P}[K = k] \propto \frac{e^{-\lambda}\lambda^k}{k!}, \quad \lambda > 0, \quad k = 1, 2, \ldots, K_{\max}, \tag{5}$$

(ii) uniform prior,

$$p_K(k) = \mathbb{P}[K = k] \propto 1, \qquad k = 1, 2, \ldots, K_{\text{max}}, \tag{6}$$

(iii) inverse prior,

$$p_K(k) = \mathbb{P}[K = k] \propto \frac{1}{k}, \qquad k = 1, 2, \ldots, K_{\text{max}}. \tag{7}$$

There are several methods to detect the presence/absence of chirp signals in the given data, e.g., see [6, 21] and [25]. So without loss of generality, we exclude 0 from the support of the distribution of $K$. Further, based on the data size, we set a reasonable choice of $K_{\text{max}}$ and any value greater than $K_{\text{max}}$ does not belong to the support of the distribution.

Truncated Poisson prior (5) is justified only when we have a good guess about the mean of $K$. Moreover, the truncated Poisson prior puts higher penalty on moderate to large number of components and hence it will tend to underestimate the model order, whereas the data may itself contain large chirp components in truth. To resolve this, we have chosen uniform prior (6) also in the comparison. We have also chosen prior (7) for comparison, which puts inverse weight to the number of components and its penalty is between that of uniform and truncated Poisson priors.

### 3.2 Prior on Amplitudes

For amplitude parameters, we consider the following prior with zero centered mean as:

$$\boldsymbol{\mathcal{A}}_K \Big| \big(K, \boldsymbol{\eta}_K, \sigma^2\big) \sim \mathcal{N}_{2K}\Big(\mathbf{0}, \delta^2\sigma^2\Big(\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big)^{-1}\Big), \quad \delta > 0, \tag{8}$$

where $\delta^2$ is taken to put more flexibility over variability of the conditional prior on the amplitude parameters.

### 3.3 Prior on Noise Variance

The prior on the noise variance is taken as Inverse Gamma distribution with scale parameter $\gamma_0/2$ and shape parameter $\nu_0/2$, i.e.,

$$\sigma^2 \sim \mathcal{IG}(\nu_0/2, \gamma_0/2). \tag{9}$$

### 3.4 Prior on Frequencies and Frequency Rates

We assume that the different components' non-linear parameters have independent priors, and for $\ell = 1, 2, \ldots, K$, the joint prior distribution $f(.)$ (probability density function) of $(\alpha_\ell, \beta_\ell)$ is uniformly distributed over the rectangle $(0, 2\pi) \times (0, \pi/2)$. Hence joint prior distribution of the non-linear parameters conditioned on $K$ is given as:

$$h\big(\boldsymbol{\eta}_K | K\big) = \prod_{\ell=1}^{K} f(\alpha_\ell, \beta_\ell). \tag{10}$$

### 3.5 Posterior Distribution and its Mode

We consider the estimation of number of components $K$ under the 0-1 loss function

$$L(K,\widetilde{K}) = \begin{cases} 0, & \widetilde{K}=K, \\ 1, & \widetilde{K}\neq K, \end{cases},$$

for which the Bayes estimator will be the argument that minimizes the posterior expected loss function

$$\mathbb{E}\Big(L(K,\widetilde{K})|\boldsymbol{Y}\Big) = \mathbb{P}[\widetilde{K}\neq K|\boldsymbol{Y}] = 1-\mathbb{P}[\widetilde{K}=K|\boldsymbol{Y}].$$

So the Bayes estimator reduces to the argument that maximizes the posterior probability mass given the data $\boldsymbol{Y}$, which is actually the mode of the posterior distribution. Therefore, we propose to use mode of the posterior distribution as the estimate of the number of components $K$. Note that 0-1 loss function is preferable over other standard loss functions like squared error loss, absolute deviation loss because the Bayes estimates then would be mean, median respectively, which can be fraction sometimes, however this is not reasonable as an estimate of a discrete random variable $K$.

Therefore we consider the mode of the posterior distribution (11) for estimation of $K$. Now for calculating the posterior, we consider the priors from (5), (8), (9) and (10), and we get the expression of joint posterior distribution of $(K,\boldsymbol{\eta}_K)$ given $\boldsymbol{Y}$ as follows:

$$\begin{aligned} p\Big(k,\boldsymbol{\eta}_k\Big|\boldsymbol{Y}\Big) &\propto \frac{\lambda^k}{(1+\delta^2)^k k!}\Big(\gamma_0+\boldsymbol{Y}^\top\mathcal{P}_k\big(\boldsymbol{\eta}_k\big)\boldsymbol{Y}\Big)^{-(N+\nu_0)/2}\prod_{\ell=1}^{k} f(\alpha_\ell,\beta_\ell) \\ &\propto \frac{\lambda^k}{(1+\delta^2)^k k!}\Big(\gamma_0+\boldsymbol{Y}^\top\mathcal{P}_k\big(\boldsymbol{\eta}_k\big)\boldsymbol{Y}\Big)^{-(N+\nu_0)/2}\left(\frac{1}{\pi^2}\right)^k, \end{aligned} \tag{11}$$

where, $\mathcal{P}_k\big(\boldsymbol{\eta}_k\big) = \boldsymbol{I}_N - \frac{\delta^2}{(1+\delta^2)}\boldsymbol{D}\big(\boldsymbol{\eta}_k\big)\Big(\boldsymbol{D}\big(\boldsymbol{\eta}_k\big)^\top\boldsymbol{D}\big(\boldsymbol{\eta}_k\big)\Big)^{-1}\boldsymbol{D}\big(\boldsymbol{\eta}_k\big)^\top$. Similarly we can derive expression of $p\Big(k,\boldsymbol{\eta}_k\Big|\boldsymbol{Y}\Big)$ for the priors (6) or (7) instead of (5). Proof of (11) can be obtained through routine steps shown in the Appendix. Note that we cannot further simplify the expression (11) to obtain marginal posterior of $K|\boldsymbol{Y}$ because the non-linear parameters cannot be eliminated by integrating the posterior (11) with respect to $\boldsymbol{\eta}_k$. One way to solve this problem is to expand the posterior (for fixed $k$) around its mode using Laplace's approximation and then integrate out the posterior with respect to $\boldsymbol{\eta}_k$. However such approximations perform poor for small sample sizes or low SNR levels. So we study the estimator of number of components based on the joint posterior available, for which we can observe the following:

$$\arg\max_{\big(k,\boldsymbol{\eta}_k\big)} p\Big(k,\boldsymbol{\eta}_k\Big|\boldsymbol{Y}\Big) = \arg\max_{\big(k,\boldsymbol{\eta}_k=\widehat{\boldsymbol{\eta}}_k\big)} p\Big(k,\boldsymbol{\eta}_k\Big|\boldsymbol{Y}\Big). \tag{12}$$

Here $\widehat{\boldsymbol{\eta}}_k$ is the minimizer of $\boldsymbol{Y}^\top\mathcal{P}_k\big(\boldsymbol{\eta}_k\big)\boldsymbol{Y}$. This implies that the argument which jointly maximizes the posterior distribution with respect to $\big(k,\boldsymbol{\eta}_k\big)$ is same as the argument that maximizes the posterior with respect to $k$ only, while the corresponding set of non-linear parameter is substituted with the minimizer of $\boldsymbol{Y}^\top \mathcal{P}_k(\boldsymbol{\eta}_k)\boldsymbol{Y}$. This relation is possible because the posterior distribution is function of $\boldsymbol{\eta}_k$ through $\boldsymbol{Y}^\top \mathcal{P}_k(\boldsymbol{\eta}_k)\boldsymbol{Y}$ only, and also the posterior is inversely related to $\boldsymbol{Y}^\top \mathcal{P}_k(\boldsymbol{\eta}_k)\boldsymbol{Y}$.

However, note that for any value of $\delta$, $\widehat{\boldsymbol{\eta}}_k$ is actually the LSE. Now, obtaining LSE is computationally challenging (see [17]), hence we use asymptotic equivalent of the LSE and computationally efficient estimator, the sequential LSE, see e.g., [13] and [17]. Therefore, from now onwards, we approximate the mode of the posterior by $\underset{\left(k,\boldsymbol{\eta}_k=\widetilde{\boldsymbol{\eta}}_k\right)}{\arg\max}\; p\left(k, \boldsymbol{\eta}_k \middle| \boldsymbol{Y}\right)$, where $\widetilde{\boldsymbol{\eta}}_k$ denotes the sequential LSE of the non-linear parameters. Our proposed estimator based on combination of sequential least squares and maximum a posteriori rule is given as:

$$\mathrm{SMAP} = \widehat{K} = \underset{\left(k,\boldsymbol{\eta}_k=\widetilde{\boldsymbol{\eta}}_k\right)}{\arg\max}\; p\left(k, \boldsymbol{\eta}_k \middle| \boldsymbol{Y}\right). \tag{13}$$

# 4 Numerical Simulations

## 4.1 Performance as SNR Varies

We study the performance of proposed estimator SMAP and compare it with that based on AIC (14), BIC (15) and the AMAP criteria (16). The AMAP rule for the model (1) has not been specifically mentioned or derived anywhere in the text, but it can be easily obtained by following steps similar to that of performed for sinusoidal model in [10] and [23]. However, one can easily obtain heuristically the penalization coefficient 10 in (16) by observing the rates of convergence of the least squares estimators of the parameters $A_k, \alpha_k$ and $\beta_k$ as $O(N^{-1/2}), O(N^{-3/2})$ and $O(N^{-5/2})$ respectively, and then using a similar argument as provided in [10]. Let us denote the residual sum of squares obtained after fitting a $k$ component chirp model to the data of size $N$ by RSS($k$), then the equivalent forms of AIC, BIC and AMAP rule as a function of $k$, the number of components has been defined as follows:

$$\mathrm{AIC}(k) = N\log(\mathrm{RSS}(k))/2 + 4k, \tag{14}$$

$$\mathrm{BIC}(k) = N\log(\mathrm{RSS}(k))/2 + 4k \times \log(N)/2, \tag{15}$$

$$\mathrm{AMAP}(k) = N\log(\mathrm{RSS}(k))/2 + 10k \times \log(N)/2. \tag{16}$$

The estimator of $K$ based on AIC, BIC and AMAP rule will be the argument that minimizes these criteria respectively.

We generate data from a two-component model (1). Further, we consider various settings of the amplitude parameters, non-linear parameters, 36 different SNR values uniformly spaced between -15 and 20, and sample size $N = 600$. We can observe from (11), that varying the values of $\nu_0$ and $\gamma_0$ will not effect the mode of the posterior distribution, so we set their values as $\nu_0 = \gamma_0 = 2$. Here, different values of $\delta^2$ actually effect mode of the posterior distribution. So, for illustration we choose two values $\delta = 1, 10$. Similarly $\lambda$ directly affects the mode of the posterior distribution, e.g., larger values of $\lambda$ will shift the mode of the posterior to larger values of the estimates of $K$. Hence, we have performed numerical simulations under sufficiently diverse values $\lambda = K_0, 1, \frac{1+K\max}{2}$, where $K_0$ represents the true number of components of the model (1) from which the data is generated. For simulated data studies, we have chosen $K_{\max} = 10$. For the above setup, we have five different posteriors from which the estimates can be compared with that of based on AIC, BIC and the AMAP criteria. These five posteriors correspond to prior on $K$ as truncated Poisson (5) with three different values of $\lambda$ and then, uniform (6) and inverse (7) prior on $K$. We further have different posteriors with respect to different values of $\delta^2$, however $\delta^2$ is not directly related with the prior information available for $K$, and usually one wants to study sole effect of different values of $\lambda$ over the estimates, and hence, we would like to get rid of $\delta^2$. One way is to put a hyperprior and integrate with respect to $\delta^2$ and obtain posterior free from $\delta^2$. This is analytically difficult to solve, because $\delta^2$ is implicitly involved in the kernel of the posterior distribution, or one can try some integral approximations, which will increase the computational difficulty and may lead to unstable estimates. So, we propose a simpler yet powerful technique to get rid of $\delta^2$, by taking minimum of the estimated components as the final estimate for each of the five posteriors. The effectiveness of this method is observed through extensive simulations. The proportion of correct estimation out of 1000 independent replications, for sample size $N = 600$ and $K_0 = 2$ is displayed in Figure 1. For higher sample sizes $N = 800, 1000$ and case $K_0 = 5$, we observe similar behaviour in the Figure 4 to 8.

We have following observations from the Figure 1:

- Numerical consistency is observed i.e., frequency of correct selection increases to 1 on increasing sample sizes and SNR;
- Proposed estimator performs better than the standard AIC, BIC and the AMAP criterion, even at low SNR or low sample sizes;
- Truncated Poisson prior (even with $\lambda$ parameter of the prior distant from true value) performs much better than the uniform and inverse kind of priors at low SNR, but use of truncated Poisson prior is encouraged only when one has confident guess for $\lambda$ parameter of the prior;

## 4.2 Performance for Close Frequency and Frequency Rates

It is natural that the performance of estimators for multi-component chirp models degrade as the closeness of the non-linear parameters of the different components increase. It would be interesting to study the behavior of the proposed estimator and compare it with that of based on AIC, BIC and the AMAP rule. It is expected that performance of the estimators based on AMAP rule should degrade faster as compared to that based on BIC, because of heavier penalty term, it underestimates the model order, and when we have too close non-linear parameters, the sum of two components tends to behave as one component easily and hence, this leads to indicate underestimation which is the severe case of difficulty for AMAP rule based method. In this section, we select sample size $N = 800$, $K_0 = 2$, $K_{\max} = 10$ and SNR level at 5 dB. Consider $(\alpha_1, \beta_1) = (1, 0.5)$ and $(\alpha_2, \beta_2) = (1 + 1/N^{\text{pow}}, 0.5 + 1/N^{1+\text{pow}})$ as the non-linear parameters of two components, then the degree of separation of non-linear parameters is governed by variable pow. As we vary variable pow from 0.1 to 2 with a gap of 0.01, we obtain deteriorating performance of all the estimators, where, the proposed estimator is clearly loosing its efficiency slowest as compared to the other three estimators based on AIC, BIC and the AMAP rule. We can further observe that estimator

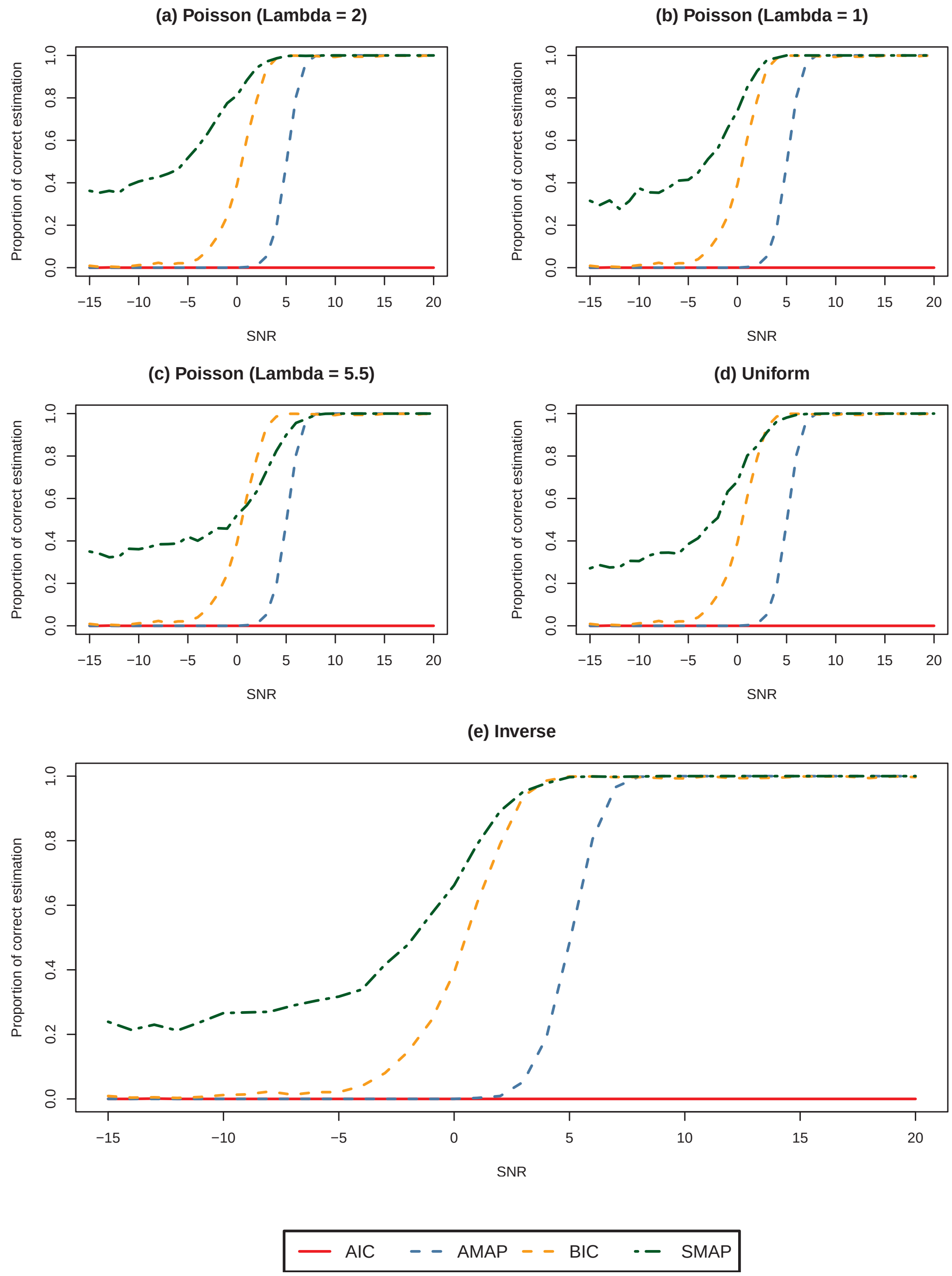



Figure 1: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 600$ is generated from a two-component chirp model. Here Poisson represents the truncated Poisson.

based on AMAP starts degrading fastest and foremost out of the four estimators, however the estimator based on BIC and the proposed estimator SMAP have either the comparable performance, or the SMAP performs better than the other two, specially when pow increases from 1 to 2.

### 4.3 Approximation to the Posterior of Number of Components

We have expression of joint posterior of $(K, \boldsymbol{\eta}_K)$, so it is desirable for many applications to evaluate the probability masses $p_{K|\boldsymbol{Y}}(k)$ over different ranges of $K$ and build Bayesian credible intervals. However, we have knowledge of $p\Big(k, \boldsymbol{\eta}_k \Big| \boldsymbol{Y}\Big)$ evaluated at sequential least squares $\boldsymbol{\eta}_k = \widetilde{\boldsymbol{\eta}}_k$ only. Another quantity of interest which can be taken as a benchmark for $p_{K|\boldsymbol{Y}}(k)$ will be $p^0_{K|\boldsymbol{Y}}(k)$ where, $p^0_K(.)$ represents the posterior distribution of $K|\boldsymbol{Y}$ obtained after using true information about the non-linear parameters $\boldsymbol{\eta}_k$ and eliminating remaining parameters by integrating them out from the joint posterior. Quantity $p^0_K(.)$ would have been ideal choice for the posterior of $K|\boldsymbol{Y}$ to build Bayesian credible intervals, however, in practice, we do not have information about $\boldsymbol{\eta}_k$ and hence, we use $p\Big(k, \boldsymbol{\eta}_k \Big| \boldsymbol{Y}\Big)$ evaluated at $\boldsymbol{\eta}_k = \widetilde{\boldsymbol{\eta}}_k$ to approximate the posterior $p_{K|\boldsymbol{Y}}(k)$ . For numerical comparison, we first generate a multi-component chirp data based on different SNR values, and we obtain various range of percentiles from these two posteriors, true posterior $p^0_{K|\boldsymbol{Y}}(k)$ and the approximated posterior $p\Big(k, \boldsymbol{\eta}_k \Big| \boldsymbol{Y}\Big)$ evaluated at $\boldsymbol{\eta}_k = \widetilde{\boldsymbol{\eta}}_k$. From the Table 1, we conclude that the approximation is pretty convincing for large range of SNR values as all the percentiles from these two posteriors either match or differ at maximum by 1 unit, when the prior on $K$ is Pois(2). We have results for different case of priors and $K_0$ in Table 1 and 5 to 13.

## 5 Real Data Analysis

For any real data, we do not have knowledge about true $K_0$, hence, we have just four posterior distributions, two corresponding to truncated Poisson with $\lambda$ parameter 1 and $\frac{1+K\text{max}}{2}$, and rest two correspond to the uniform and inverse prior case. We have assumed distribution of error term $X(n)$ to be i.i.d. normal with mean 0 and variance $\sigma^2$. Hence, it will be reasonable to utilize tests of normality for the residuals obtained corresponding to these four estimates.

So, we choose the minimum of these four components estimates, whose corresponding residuals pass the i.i.d. normality tests. We have chosen Anderson-Darling test, Shapiro-Wilk Normality test, Jarque–Bera test, Cramer-von Mises test, Lilliefors (Kolmogorov-Smirnov) test, Pearson chi-square test and Shapiro-Francia test for normality. We say that a residual data passes the normality test if none of these seven tests detect significant non-normality behavior. Here, we analyze two different applications based real datasets, first is bat data and second is electroencephalogram (EEG) data. Bat data has available sample size 400. We utilize only initial 300 sample points to fit the model (1) and obtain the estimates based on proposed method SMAP, AIC, BIC and the AMAP rule. We have left rest of the new points from 301 to 400, to assess the performance of the fitted model based on prediction error sum of squares. We compare the estimates in Table 2 and prediction error sum of squared errors corresponding to different estimators in the Table 3.

We also conclude a reasonable fit of the proposed estimator which is neither based on a over-simplified

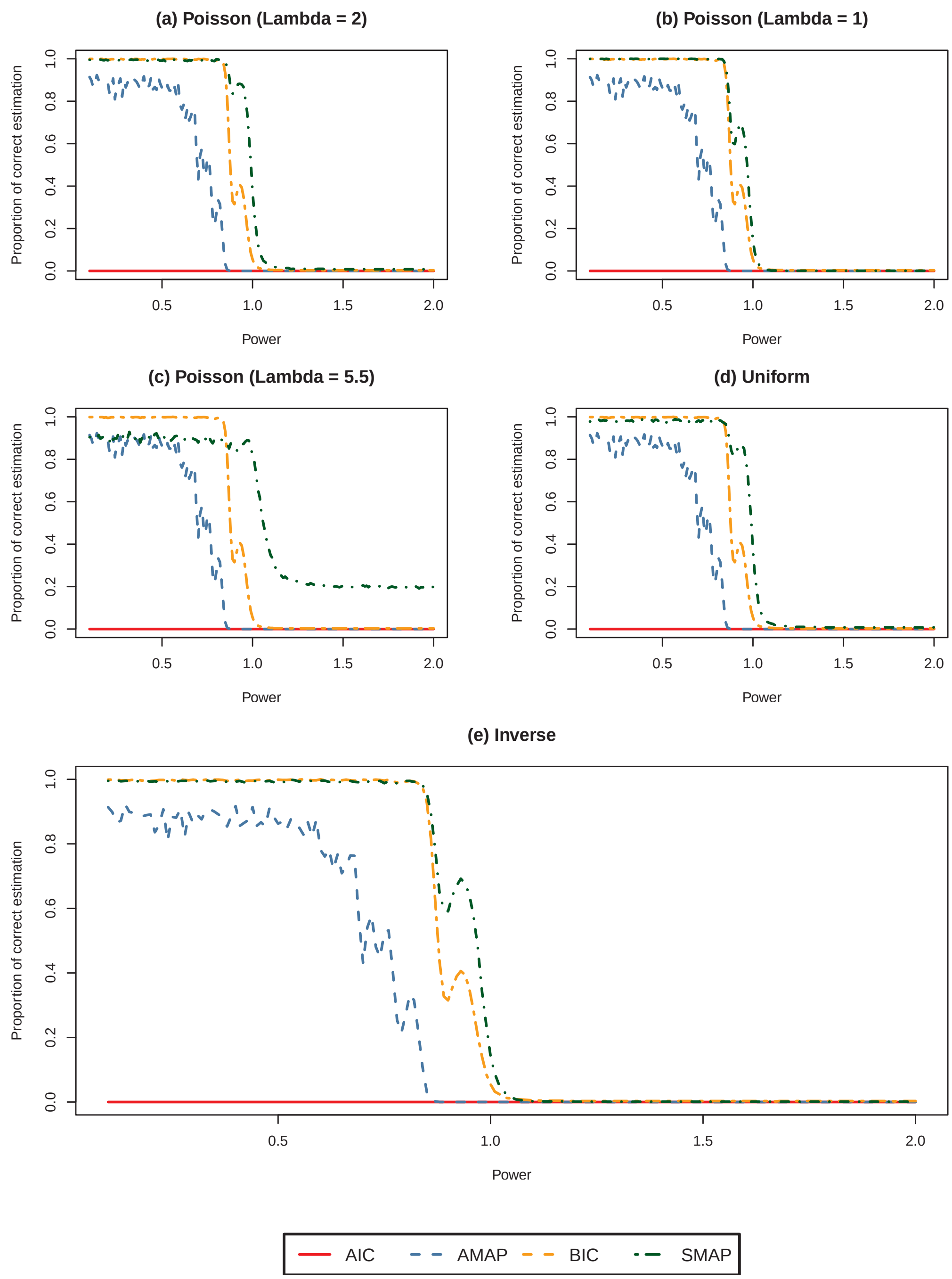



Figure 2: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as power pow varies when the data is generated from a two-component chirp model. Here Poisson represents the truncated Poisson.

| SNR | Posterior | | | | | Percentiles | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 10 dB | True | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 20 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 18 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 25 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 30 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |

Table 1: Quantiles for the True Posterior Distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior Pois(2) and $\boldsymbol{Y}$ is obtained from a two-component chirp model at sample size $N = 800$.

model (as AMAP performs) nor over-complicated model (as AIC and BIC performed for these data). Further, we observe that our proposed estimator had lower prediction error sum of squares than that of the over-complicated model obtained by using AIC or BIC. Component estimate 3 and 50 do not pass the normality test for bat data and hence, we obtain 5 as the proposed estimate of $K$.

Similarly, we observe results for the EEG data, where we obtain estimates for initial 200 size data and then based on data points 201 to 256, we obtain prediction error sum of squares. From Table 2, we have 4 as our proposed estimate for the EEG data, as residuals corresponding to 4 passes the normality test. However, residuals corresponding to 1 and 20 components fit, do not pass the normality test. The AIC and BIC based estimators for both the datasets select an over-complicated model, while AMAP rule selects an over-simplified model, e.g., see Figure 3b, which is the reason for their corresponding residuals behaving significantly non-normal. Further note that the residuals corresponding to AIC, BIC and the AMAP rule do not pass the normality tests for both datasets, whereas this is not the case with our proposed estimator. Therefore, SMAP provides a reasonable fit with comparatively less prediction error sum of squares than the that of AIC and BIC. We further add the comparison of 90 percentile values for each of the priors based on these two datasets. From Table 4, one can easily conclude that the AIC and BIC based estimates are less probable than those of the proposed method. We further point out that the estimates of $K$ based on AMAP criterion seems to be more reasonable than those of based on AIC and BIC, however the fit based on AMAP rule for EEG data is over simplified.

| Data | $K_{\max}$ | AIC | BIC | AMAP | Proposed Estimator SMAP | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Pois(1) | Pois$\left(\frac{1+K_{\max}}{2}\right)$ | Uniform | Inverse |
| Bat | 50 | 50 | 50 | 3 | 3 | 6 | 5 | 5 |
| EEG | 20 | 20 | 20 | 1 | 4 | 6 | 4 | 4 |

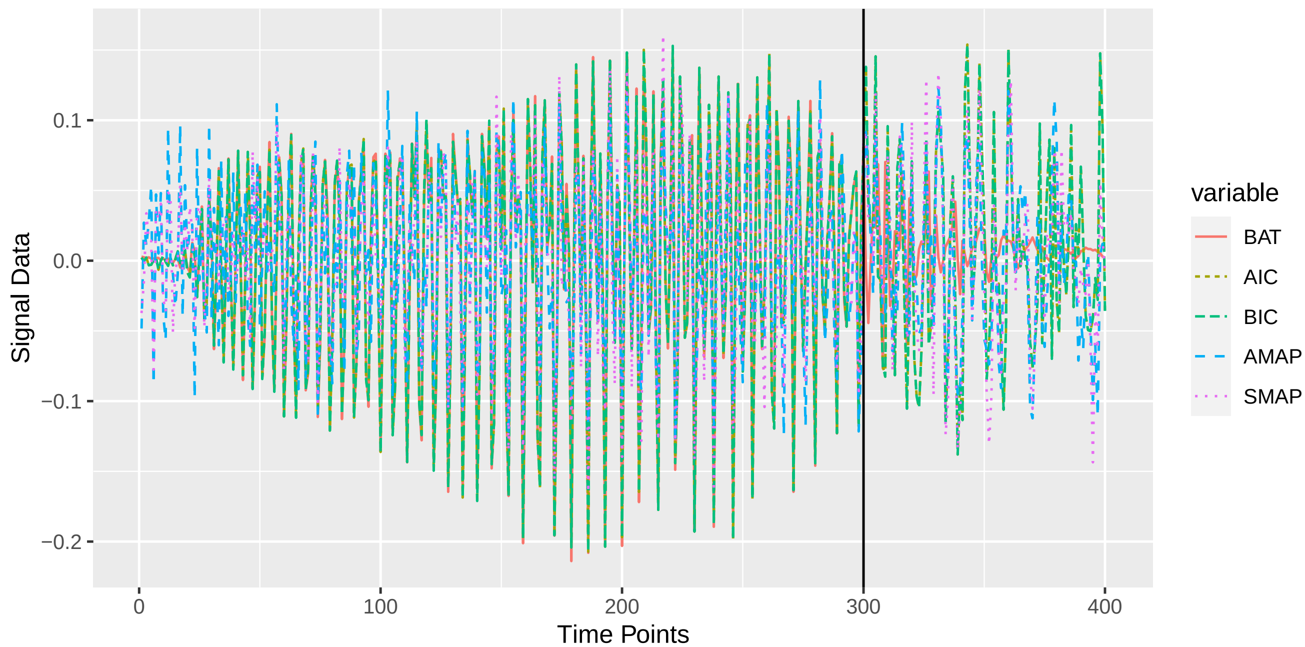
Table 2: Estimates of $K$ based on real datasets EEG and bat data.

| Data | AIC | BIC | AMAP | SMAP |
|---|---|---|---|---|
| Bat | 0.0050 | 0.0050 | 0.0036 | 0.0042 |
| EEG | 55.3547 | 55.3547 | 20.6847 | 39.9258 |

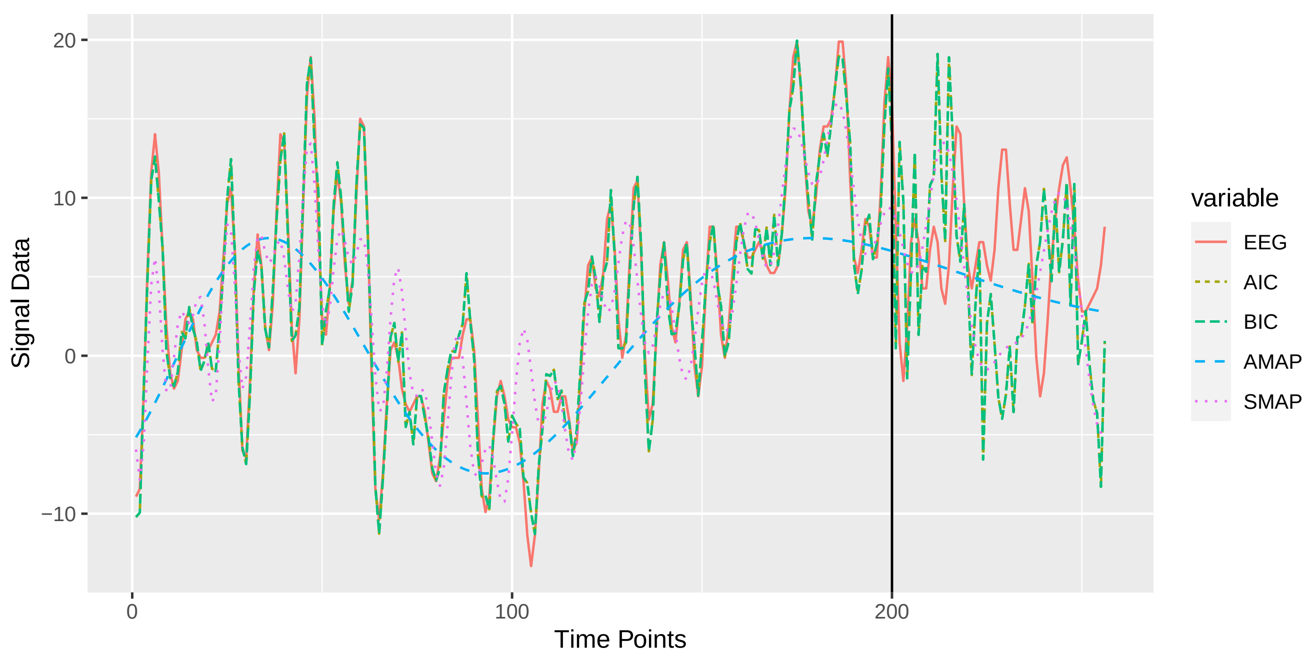
Table 3: Predicted error sum of squares obtained on fitting a multi-component chirp model with estimated model order.

| Percentile | Prior | Bat Data | | | | EEG data | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pois(1) | Pois(25.5) | Unif | Inverse | Pois(1) | Pois(10.5) | Unif | Inverse |
| 90 % | delta 1 | 38 | 43 | 43 | 43 | 18 | 18 | 18 | 18 |
| | delta 10 | 33 | 38 | 38 | 38 | 19 | 19 | 19 | 19 |
| 80 % | delta 1 | 31 | 36 | 36 | 36 | 15 | 16 | 16 | 16 |
| | delta 10 | 25 | 31 | 30 | 30 | 18 | 18 | 18 | 18 |

Table 4: 90% and 80% quantiles of the number of components based on different posterior distributions.




(a) Real bat data fitted by multi-component chirp model.




(b) Real EEG data fitted by multi-component chirp model.

Figure 3: Estimated model orders based on AIC, BIC, AMAP and SMAP are used to obtain fit for the real datasets, where data before the black vertical line represents the data used to fit the model parameters, whereas the data after black line represents the new data points or prediction points.

## 6 Conclusion

We have proposed an estimator of the number of components based on approximating the mode of the posterior distribution. Through extensive simulations, we demonstrate the superiority of the proposed estimator over that of the standard estimators based on minimizing the AIC, BIC and AMAP criteria. Reasonable performance of the proposed estimator is further validated through detailed analysis of two real datasets. The AIC and BIC based methods are overestimating the model order while AMAP based method is underestimating a lot, for the considered multi-component chirp model. We have proposed SMAP which provides reasonable balance between the complexity of the model and flexibility to predict new data points. The proposed method can easily be generalized for order estimation of multi-component models with forms general than chirps. This work motivates many problems, e.g., including data-dependent prior for the amplitudes, utilizing Bayesian least absolute shrinkage and selection operator (LASSO) for estimating the number of chirp components, etc.

# 7 Appendix

**Proof of** (11)**:**

*Proof.*

$$
\begin{aligned}
&p\Big(K, \boldsymbol{\eta}_K, \boldsymbol{\mathcal{A}}_K, \sigma^2 \Big| \boldsymbol{Y}\Big) \propto \\
&\Big(\frac{1}{\sigma^2}\Big)^{N/2} \exp\Big( - \frac{1}{2\sigma^2}\Big(\boldsymbol{Y} - \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\boldsymbol{\mathcal{A}}_K\Big)^\top \Big(\boldsymbol{Y} - \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\boldsymbol{\mathcal{A}}_K\Big)\Big) \prod_{\ell=1}^{K} f(\alpha_\ell, \beta_\ell) \\
&\times \left| \frac{\Big(\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big)}{2\pi\sigma^2\delta^2} \right|^{1/2} \exp\Big( - \frac{1}{2\delta^2\sigma^2}\Big(\boldsymbol{\mathcal{A}}_K^\top \Big(\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big)\boldsymbol{\mathcal{A}}_K\Big)\Big) \\
&\times \frac{\lambda^K}{K!} \exp\Big( - \frac{\gamma_0}{2\sigma^2}\Big)\Big(\frac{1}{\sigma^2}\Big)^{\nu_0/2+1} \\
&\propto \Big(\frac{1}{\sigma^2}\Big)^{(\nu_0+N)/2+1} \frac{\lambda^K}{K!} \exp\Big( - \frac{1}{2\sigma^2}\Big(\gamma_0 + \boldsymbol{Y}^\top \mathcal{P}_K\big(\boldsymbol{\eta}_K\big)\boldsymbol{Y}\Big)\Big) \prod_{\ell=1}^{K} f(\alpha_\ell, \beta_\ell) \\
&\times \left| \frac{1}{2\pi\sigma^2 \mathcal{M}_K (1+\delta^2)} \right|^{1/2} \exp\Big( - \frac{1}{2\sigma^2}\Big(\boldsymbol{\mathcal{A}}_K - \boldsymbol{m}_K\Big)^\top \mathcal{M}_K^{-1}\Big(\boldsymbol{\mathcal{A}}_K - \boldsymbol{m}_K\Big)\Big),
\end{aligned}
$$

where, $\mathcal{M}_K^{-1} = \big(1 + \frac{1}{\delta^2}\big)\Big(\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big)$, $\boldsymbol{m}_K = \mathcal{M}_K \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{Y}$ and $\mathcal{P}_K\big(\boldsymbol{\eta}_K\big) = \boldsymbol{I}_N - \dfrac{\delta^2}{(1+\delta^2)} \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big(\boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)\Big)^{-1} \boldsymbol{D}\big(\boldsymbol{\eta}_K\big)^\top$. Integrating with respect to $\boldsymbol{\mathcal{A}}_K$ and $\sigma^2$ yields (11).

$\square$

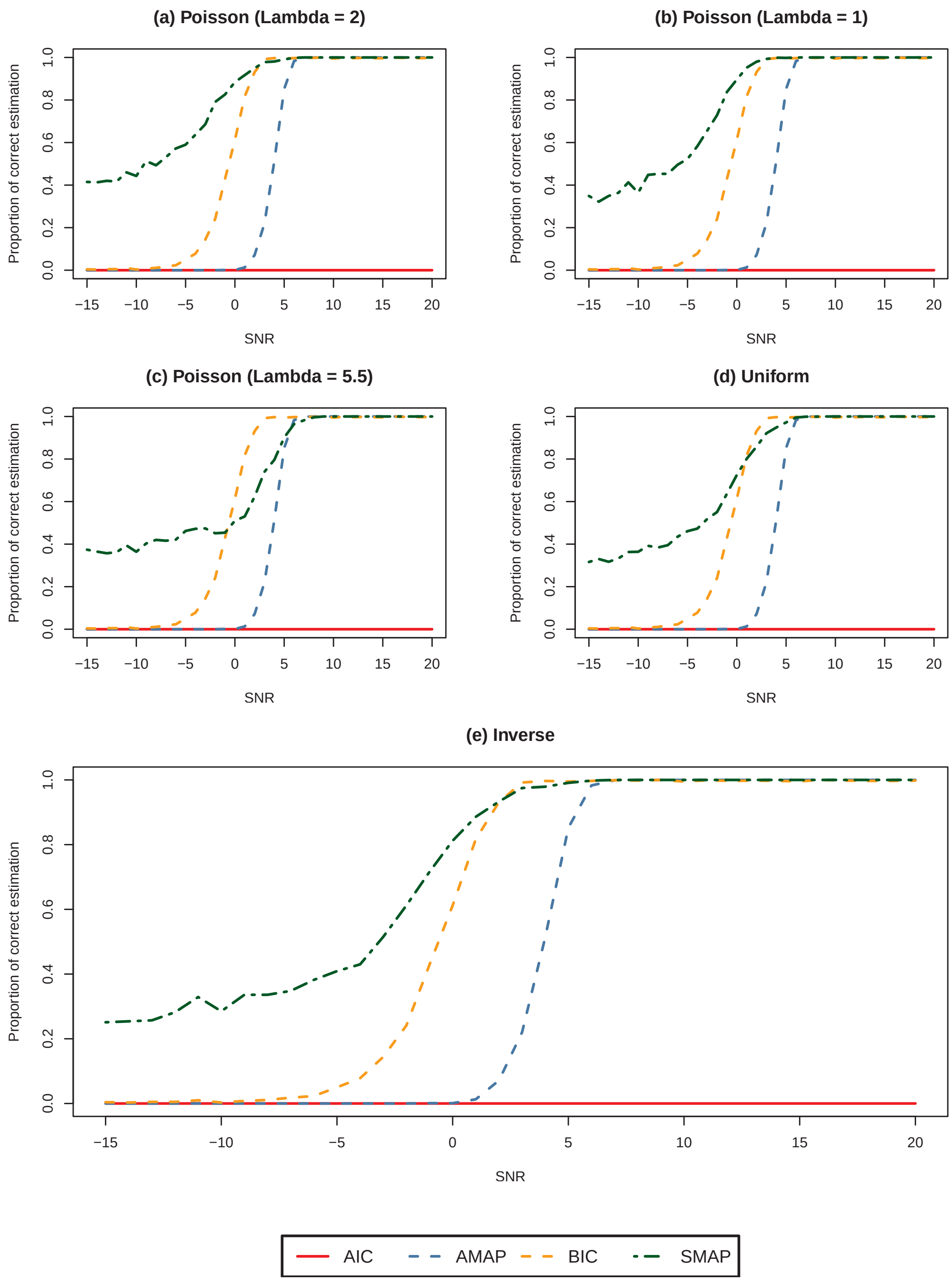



Figure 4: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 800$ is generated from a two-component chirp model. Here Poisson represents the truncated Poisson.

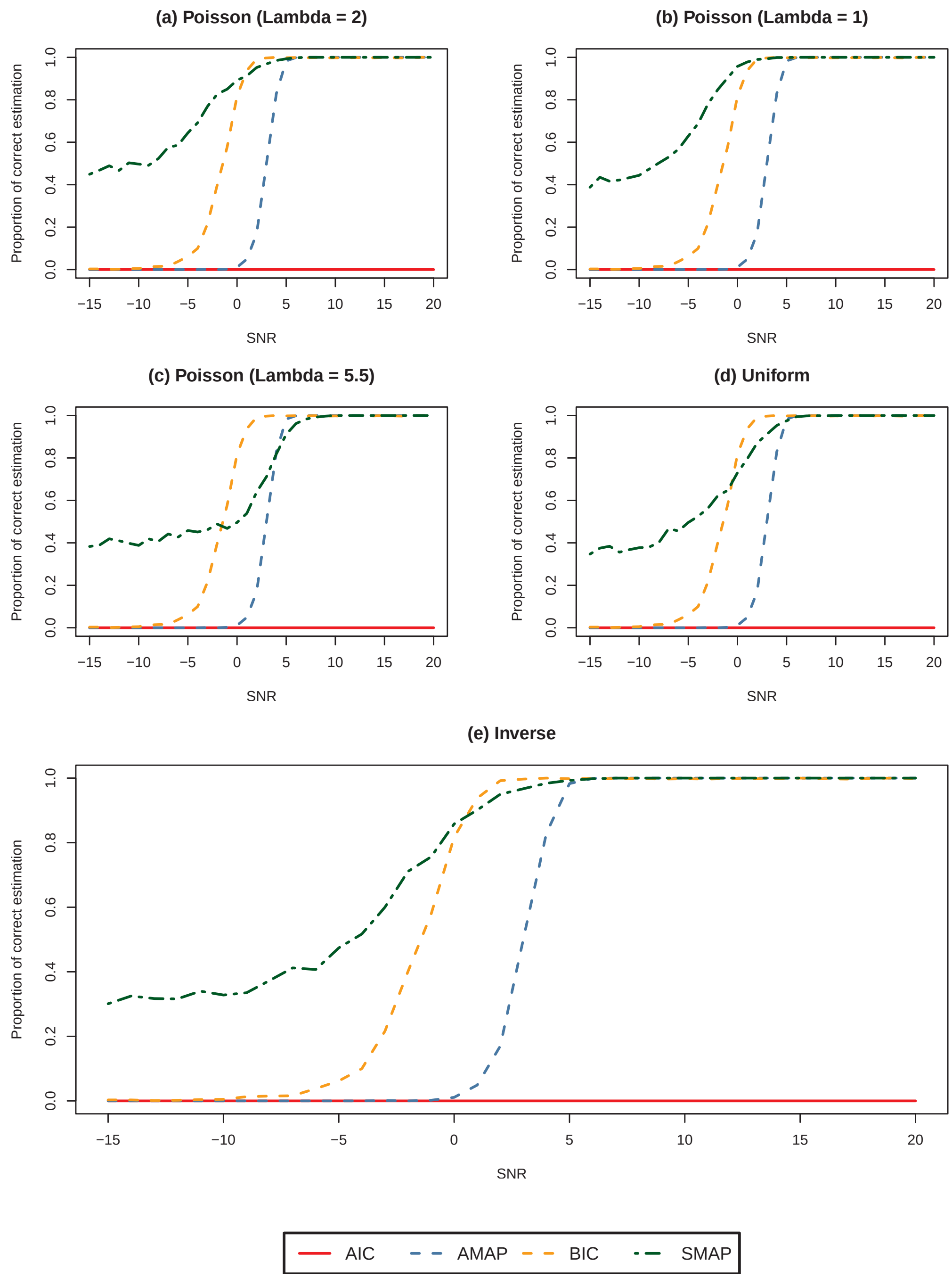



Figure 5: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 1000$ is generated from a two-component chirp model. Here Poisson represents the truncated Poisson.

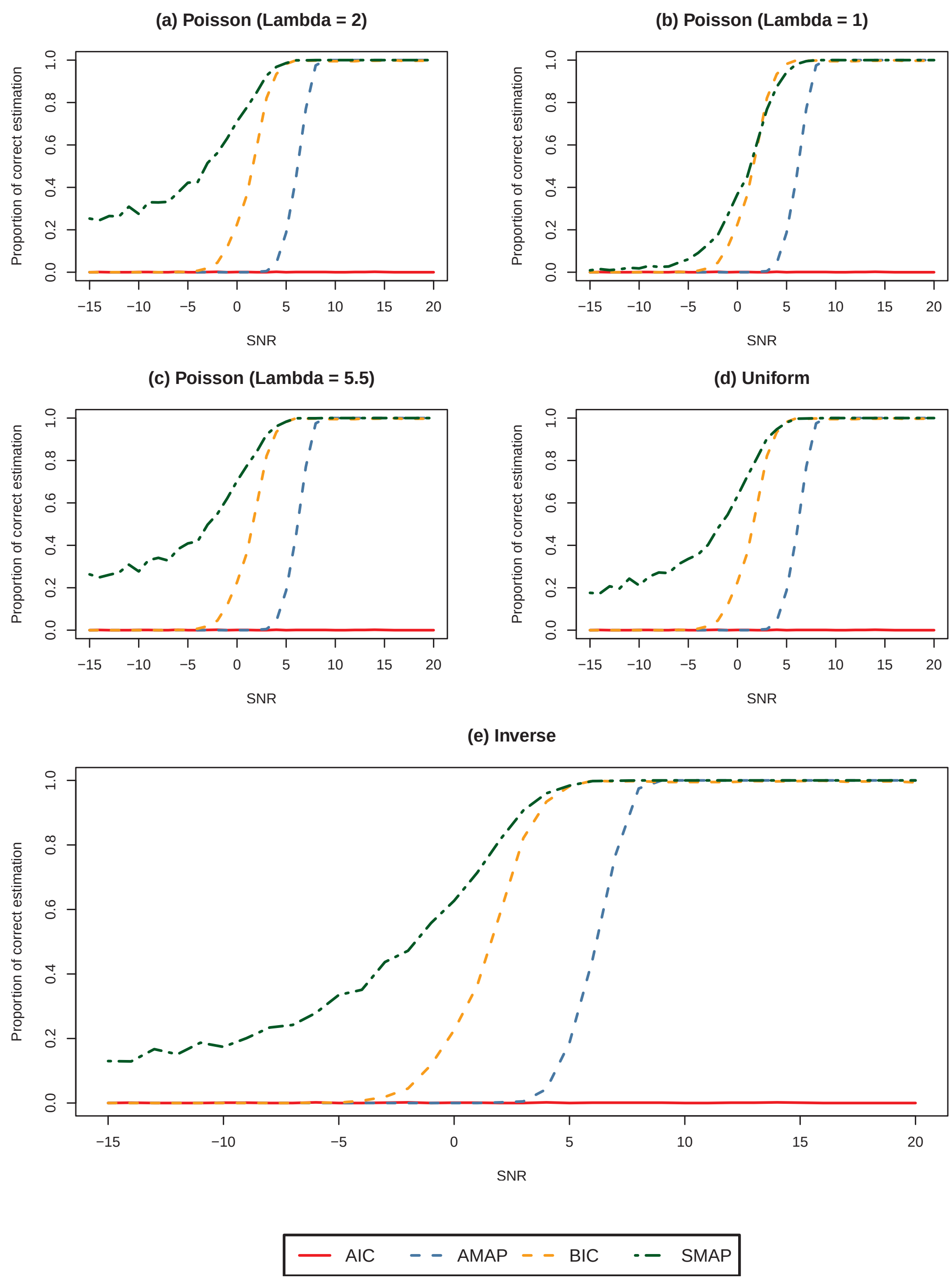



Figure 6: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 600$ is generated from a five-component chirp model. Here Poisson represents the truncated Poisson.

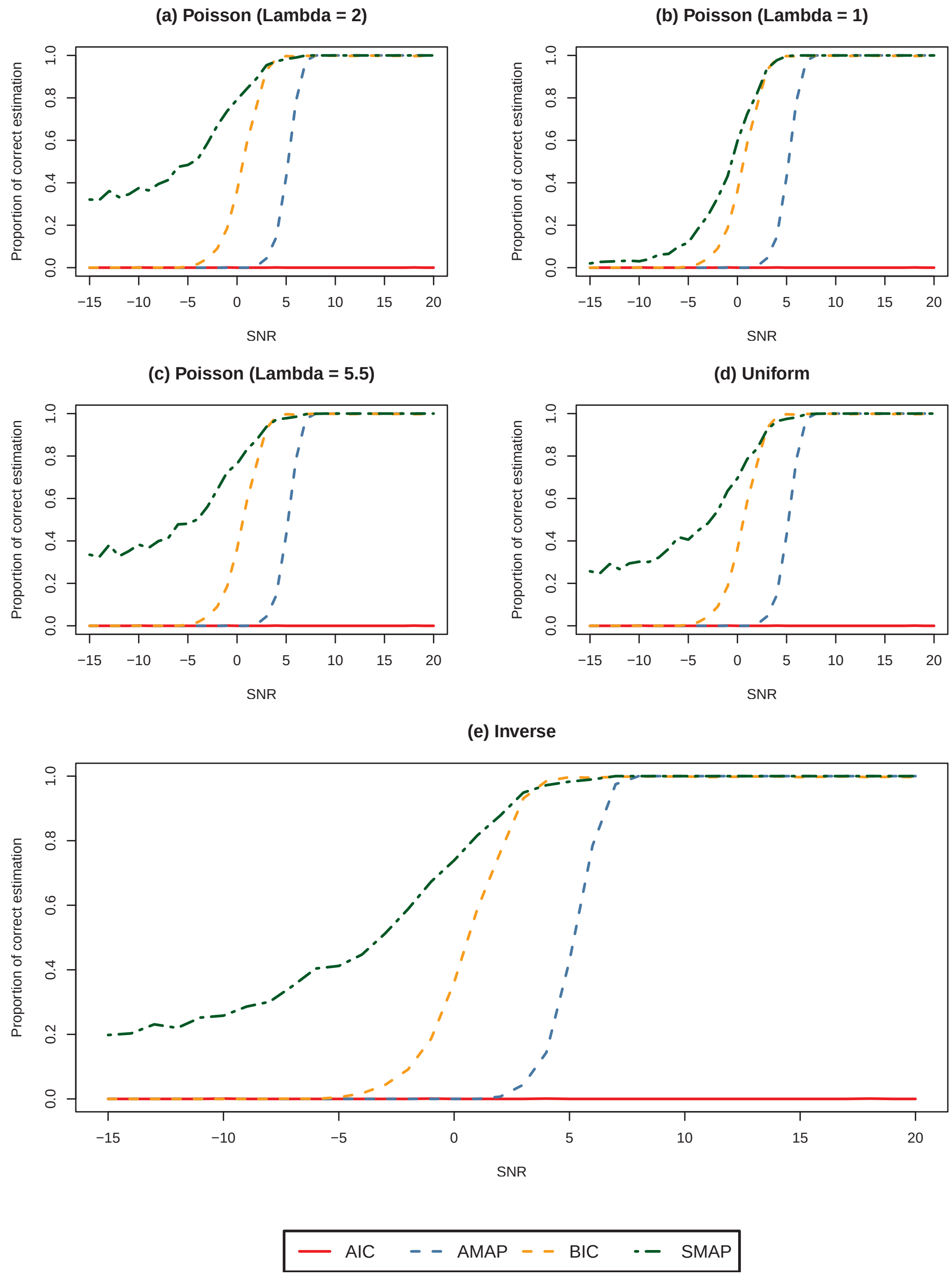



Figure 7: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 800$ is generated from a five-component chirp model. Here Poisson represents the truncated Poisson.

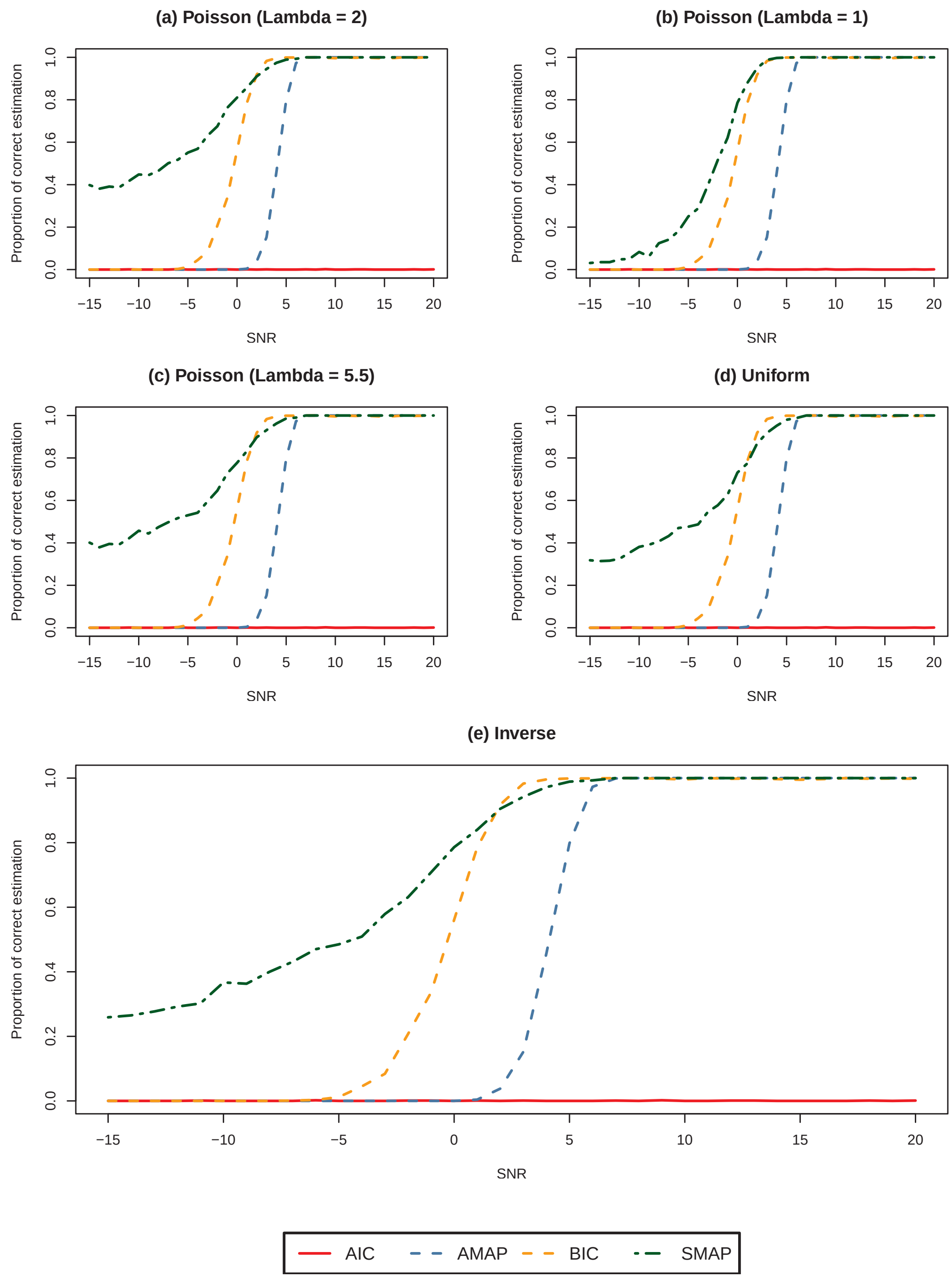



Figure 8: Performance of AIC, AMAP, BIC and proposed estimator SMAP, as SNR varies when the data of sample size $N = 1000$ is generated from a five-component chirp model. Here Poisson represents the truncated Poisson.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 10 dB | True | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 15 dB | True | 3 | 5 | 7 | 9 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 20 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 25 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 16 | 18 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 30 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 18 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |

Table 5: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior Pois(1) and $\boldsymbol{Y}$ is obtained from a two-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 5 | 7 | 9 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 18 |
| | Approximated | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 10 | 12 | 14 | 16 | 18 |
| 15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| 20 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| 25 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 30 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |

Table 6: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior truncated Pois$\left(\frac{1+K\text{max}}{2}\right)$ and $\boldsymbol{Y}$ is obtained from a two-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 15 | 16 | 18 |
| 15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| 20 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| 25 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 30 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |

Table 7: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the uniform prior and $\boldsymbol{Y}$ is obtained from a two-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| 20 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 10 | 12 | 14 | 16 | 18 |
| 25 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 30 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |

Table 8: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the inverse prior and $\boldsymbol{Y}$ is obtained from a two-component chirp model at sample size $N = 800$.

| SNR | Posterior | | | | | Percentiles | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| 10 dB | True | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| 15 dB | True | 5 | 7 | 8 | 10 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| 20 dB | True | 6 | 7 | 9 | 11 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 25 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |
| 30 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |

Table 9: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior Pois(5) and $\boldsymbol{Y}$ is obtained from a five-component chirp model at sample size $N = 800$.

| SNR | Posterior | | | | | Percentiles | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -5 dB | True | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| 10 dB | True | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 4 | 6 | 8 | 9 | 11 | 13 | 15 | 17 | 19 |
| 15 dB | True | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| | Approximated | 5 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| 20 dB | True | 6 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| 25 dB | True | 6 | 7 | 9 | 11 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |
| 30 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |

Table 10: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior Pois(1) and $\boldsymbol{Y}$ is obtained from a five-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 15 | 17 | 19 |
| -5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| 10 dB | True | 4 | 6 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| | Approximated | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| 15 dB | True | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| 20 dB | True | 6 | 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 25 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 30 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |

Table 11: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the prior truncated Pois$\left(\frac{1+K_{\max}}{2}\right)$ and $\boldsymbol{Y}$ is obtained from a five-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 15 | 17 | 19 |
| -5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| 10 dB | True | 4 | 6 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| | Approximated | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| 15 dB | True | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| 20 dB | True | 6 | 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 25 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 30 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |

Table 12: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the uniform prior and $\boldsymbol{Y}$ is obtained from a five-component chirp model at sample size $N = 800$.

| SNR | Posterior | Percentiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| -15 dB | True | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| -10 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| | Approximated | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 17 | 19 |
| -5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| 0 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 14 | 16 | 18 |
| 5 dB | True | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| | Approximated | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| 10 dB | True | 4 | 6 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| | Approximated | 4 | 6 | 8 | 10 | 11 | 13 | 15 | 17 | 19 |
| 15 dB | True | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| | Approximated | 5 | 7 | 8 | 10 | 12 | 13 | 15 | 17 | 19 |
| 20 dB | True | 6 | 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 5 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 25 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 14 | 15 | 17 | 19 |
| 30 dB | True | 6 | 8 | 9 | 11 | 12 | 14 | 16 | 17 | 19 |
| | Approximated | 6 | 7 | 9 | 10 | 12 | 13 | 15 | 17 | 19 |

Table 13: Quantiles for the true posterior distribution $p^0_{K|\boldsymbol{Y}}(k)$ when both linear and non-linear parameters are assumed to be known, and the approximated posterior for the inverse prior and $\boldsymbol{Y}$ is obtained from a five-component chirp model at sample size $N = 800$.

# References


[1] D Anderson and K Burnham. Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020):10, 2004.

[2] Tomohiro Ando. *Bayesian model selection and statistical modeling*. CRC Press, Boca Raton, Florida, 2010. ISBN 978-1-4398-3614-9.

[3] Christophe Andrieu and Arnaud Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing*, 47(10):2667–2676, 1999.

[4] Sergio Barbarossa, Anna Scaglione, and Georgios B Giannakis. Product high-order ambiguity function for multicomponent polynomial-phase signal modeling. *IEEE Transactions on Signal Processing*, 46(3):691–708, 1998.

[5] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

[6] Rong Chen and Yiming Wang. Efficient detection of chirp signals based on the fourth-order origin moment of fractional spectrum. *Circuits, Systems, and Signal Processing*, 33:1585–1596, 2014.

[7] Gerda Claeskens, Nils Lid Hjort, et al. Model selection and model averaging. *Cambridge Books*, 2008.

[8] Manuel Davy, Christian Doncarli, and J-Y Tourneret. Hierarchical Bayesian learning using reversible jumps MCMC for chirps classification. *U.K., Technical Report CUED/F-INFENG/TR.422*, 2001.

[9] Manuel Davy, Christian Doncarli, and J-Y Tourneret. Classification of chirp signals using hierarchical Bayesian learning and MCMC methods. *IEEE Transactions on Signal Processing*, 50(2):377–388, 2002.

[10] Petar M Djuric. A model selection rule for sinusoids in white Gaussian noise. *IEEE Transactions on Signal Processing*, 44(7):1744–1751, 1996.

[11] Víctor Elvira and Emilie Chouzenoux. Optimized population Monte Carlo. *IEEE Transactions on Signal Processing*, 70:2489–2501, 2022.

[12] Peter J Green and David I Hastie. Reversible jump MCMC. *Genetics*, 155(3):1391–1403, 2009.

[13] Rhythm Grover, Debasis Kundu, and Amit Mitra. On approximate least squares estimators of parameters of one-dimensional chirp signal. *Statistics*, 52(5):1060–1085, 2018.

[14] Mingyi Hong, Mónica F Bugallo, and Petar M Djuric. Joint model selection and parameter estimation by population Monte Carlo simulation. *IEEE Journal of Selected Topics in Signal Processing*, 4(3): 526–539, 2010.

[15] Muhammad Z Ikram and G Tong Zhou. Order determination and parameter estimation of multicomponent polynomial phase signals. In *Ninth IEEE Signal Processing Workshop on Statistical Signal and Array Processing (Cat. No. 98TH8381)*, pages 352–355. IEEE, 1998.

[16] R Kashyap. Inconsistency of the aic rule for estimating the order of autoregressive models. *IEEE Transactions on Automatic Control*, 25(5):996–998, 1980.

[17] Ananya Lahiri, Debasis Kundu, and Amit Mitra. Estimating the parameters of multiple chirp signals. *Journal of Multivariate Analysis*, 139:189–206, 2015. doi: 10.1016/j.jmva.2015.01.019.

[18] Chung-Chieh Lin and Petar M Djuric. Estimation of chirp signals by MCMC. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 1, pages 265–268. IEEE, 2000.

[19] Heinz Linhart and Walter Zucchini. *Model selection.* John Wiley & Sons, United Stalcs of America, 1986. ISBN 0-471-83722-9.

[20] Allan DR McQuarrie and Chih-Ling Tsai. *Regression and time series model selection.* World Scientific, Singapore, 1998. ISBN 981-02-3242-X.

[21] Mohamed Abou Bakr Othman, John Belz, and Behrouz Farhang-Boroujeny. Performance analysis of matched filter bank for detection of linear frequency modulated chirp signals. *IEEE Transactions on Aerospace and Electronic Systems*, 53(1):41–54, 2017.

[22] Bingxin Shen, Mónica F Bugallo, and Petar M Djurić. Estimation of multimodal posterior distributions of chirp parameters with population Monte Carlo sampling. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3861–3864. IEEE, 2012.

[23] Petre Stoica and Yngve Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[24] Khushboo Surana, Sharmishtha Mitra, Amit Mitra, and Petre Stoica. Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties. *Signal Processing*, 128:204–211, 2016.

[25] Bo Tang, Haibo He, and Steven Kay. Adaptive signal detection and parameter estimation in unknown colored Gaussian noise. *arXiv preprint arXiv:1607.08259*, 2016.

[26] Peng Yang, Zheng Liu, and Wen-Li Jiang. Parameter estimation of multi-component chirp signals based on discrete chirp fourier transform and population Monte Carlo. *Signal, Image and Video Processing*, 9:1137–1149, 2015.

[27] Peng Yang, Hong Ding, Hui Xiong, and Linhua Zheng. Adaptive importance sampling for estimating multi-component chirp signal parameters in colored noise. *Signal, Image and Video Processing*, 13: 617–625, 2019.