RATIONAL SPECTRAL ESTIMATION BY REVERSIBLE JUMP MCMC

JARED A. MCBRIDE

Abstract. In this report I apply a reversible jump Markov chain Monte Carlo technique to the problem of fitting an autoregressive (AR) model to data. In doing this I focus on the proposed poles of the AR model. A review of relevant background material is included. And a full specification of the implementation is provided. At this time the attempt has been unsuccessful and only partial results are reported.

Contents

1. Introduction

Myriad fields of industry and research are tasked with the analysis of timeseries of one form or another. Many of these times series fall into the category of stationarity and therefore may be studied by power spectrum analysis. Such analysis has proven to be a useful and effective tools in these fields, particularly in estimation and control theory, to name a few. Usually the power spectrum must be estimated from data.

The present work serves as a stepping stone to the task of rational spectral approximation given signal observations. Rational approximations can be preferred over the common Laurent polynomial approximation since they often can achieve great accuracy with much fewer parameters. The task of Rational approximation is essentially that of ARMA fitting. And the hope of this work is to provide guidance for solving that problem by first considering AR fitting by reversible jump Markov chain Monte Carlo over the space of poles and error variance, which things determine the centered AR process. In a sense this is a proof of

Date: May 16, 2022.

concept or trial run to understand the problem well enough to better determine whether the full ARMA problem is worthwhile.

Spectral factorization is a very common operation preformed on estimated power spectrum. The most common algorithms for spectral factorization assume the spectrum is in the form of a Laurent polynomial [1, 2], though recently some methods have been developed in greater generality [3, 4]. If positivity on the unit circle is enforced a rational power spectrum estimate can be written as the quotient of Laurent polynomials and thus rational power spectrum estimates are amenable to the most common spectral factorization algorithms. In their survey Kailath et al. discussed a number of factorization techniques, all of which were iterative and had a run time of at least $O(N^2)$ per iteration, here N is the degree of the Laurent polynomial. And so, in some cases fitting a rational prior to factoring will afford a savings in computational cost since, as was mentioned above, comparable if not great accuracy can be achieved with many fewer degrees.

Though there are other more direct methods to estimate ARMA models such as maximal likelihood estimates and Burg estimates (see, for example [5]) these techniques require the user to select the order of the ARMA, *a priori*, or compute analysis on a range of orders and identify a criteria (such as AIC or BIC) to select the best order (for an example see [6]). In contrast this method has no such constraint. The order is left as a parameter to be fitted with the rest.

I have decided to focus on the poles and zeros of the characteristic functions of the ARMA model (or for now the AR model) for a few reasons. One is that ARMA and AR models require their poles to lie inside the unit circle in order for the process to be stationary. To propose poles already in the unit circle is much easier to work with than the analogous constraint applied to the autoregressive coefficients. Another reason to focus on the zeros and poles is that there is a uniform sensitivity among the them. The autoregressive coefficients are ordered and each plays a unique role in the estimator, the coefficient to the first lagged term has a different job than that on the 10th lagged term, but there is no ordering to the poles and they enjoy homogeneity among themselves. Though this has not been substantiated, I believe this may make the space more regular, and them problem better conditioned.

The remainder of the report is organized as follows. In Section 2, I provide relevant backgrounds material including some results from autoregressive models (Sec 2.1), an introduction to reversible jump Markov chain Monte Carlo (RJMCMC) (Sec 2.2), and Bayesian inference in the framework of Markov chain Monte Carlo (MCMC) (Sec 2.3). I do assume the reader has some familiarity with basic MCMC theory. Then in Section 3, I construct the RJMCMC implementation that will be used to explore the space of perspective AR models. In Section 4, I briefly outline the experiments I ran and summarize what results were obtained. And finally in Section 5, I conclude with some remarks.

2. Background

In this section I provide material useful to understanding what is to follow. I first define the autoregressive model, and provide some relevant properties, including a derivation of the likelihood of an AR process given model parameters. I next review Bayesian inference and

how it is used in conjunction with Markov chain Monte Carlo. Following that I include an introduction to reversible jump Markov chain Monte Carlo. This includes a dirivation of the acceptance probabilities for moves from spaces of different dimension. The material of the discussion is drawn from three main sources [7, 8, 9].

2.1. **The Autoregressive Model.** There are a number of ways to model sequential data, that is, data observed regularly over time, timeseries. The autoregressive model is one such model. It assumes that the present value of the process is a random IID normal perturbation of a fixed linear combination of previous values. Written precisely, the autoregressive model of order p (AR(p)) has the form

$$Y_n + a_1 Y_{n-1} + \cdots + a_p Y_{n-p} = \mu + \varepsilon_n \qquad \varepsilon_n \sim N(0, \sigma^2) \quad \text{IID}$$

Observe that an AR(1) process is Markov.

With that definition in hand we now review some related concepts. a stochastic process $X = (X_t,\ t > -\infty)$ is said to be wide-sense stationary if (1) its mean (taken at each time)

$$\mathbb{E}X_t = \mu_t$$

does not depend on time, that is, it is constant ($\mu_t = \mu$) and (2) if autocovariance function

$$C_{XX}(t+\tau, t) = \mathbb{E}[(X_{t+\tau} - \mu)(X_t - \mu)*]$$

depends only on the lag τ in which case I denote it $C_X[\tau]$, the brackets following a signals and systems engineering convention, in discrete time.

And so, given a wide-sense stationary (WSS) stochastic process $X = \big(X_n, n = 0, 1, \dots\big)$ it's z-spectrum $S_X(z)$ is defined by

$$S_X(z) = \sum_{n=-\infty}^{\infty} C_X[n] z^{-n}.$$

Where as the z-spectrum is the z-transform[1] of the of the autocovariance sequence the term power spectrum will be used to denote the Fourier transform of the autocovariance sequence,

$$\bar{S}_X(\omega) = \sum_{n=-\infty}^{\infty} C_X[n] e^{-jn\omega} \qquad \Big(= S_X(e^{i\omega}) \Big).$$

The function $S_X(z)$ is real-valued on the unit circle since, $C_X[-n] = C_X^*[n]$, and by the Wiener-Khinchin theorem, it is, in fact, nonnegative there as well.

Returning to the AR model, to name the full random process I will omit the time index and write $X = (X_n,\ n > -\infty)$, and for the disturbance or error process $\varepsilon = (\varepsilon_n,\ n > -\infty)$. For the vector of coefficients (including $a_0 = 1$) I will write $a = (1, a_1, a_2, \dots, a_p)$

[1]The z-transform of an ℓ^1 sequence $(u_j,\ j > -\infty)$ is the function over the complex plan given by

$$\sum_{j=-\infty}^{\infty} u_j z^{-j}.$$

For the analysis to follow we will only consider data from processes that are assumed to be WSS and it will make no difference what the mean μ of the process is, so it will be set to $\mu = 0$. Thus I will only consider *centered* models and data.

An AR(p) process with coefficients a is WSS stationary if the zeros of the z-transform of a, $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}$, which I call the transfer function of the model, lie strictly within the unit circle (see e.g. [10]). With this in mind, given a WSS AR(p) process it's z-spectrum can be directly computed using the model data a and σ^2. Observe that since[2]

$$(a \star Y)_n = \varepsilon_n, \tag{1}$$

it can be shown[3] that

$$S_Y(z) = \frac{\sigma^2}{A(z)A^*(z^{-*})} = \frac{\sigma^2}{(1 - z_1 z^{-1})(1 - z_1^* z) \cdots (1 - z_p z^{-1})(1 - z_p^* z)}$$

where z_j for $j = 1, \ldots, p$ are the zeros of $A(z)$. This demonstrates that the z-spectrum of the (WSS) AR(p) process is determined by it's poles ($z_j,\ j = 1, \ldots, p$) and error covariance σ^2.

2.1.1. *The exact likelihood of a finite time realization of an AR(p) process.* I now derive the likelihood formula for a finite time truncation, $\tilde{Y} = (\tilde{Y}_n,\ n = 1, 2, \ldots, M)$ of the complete AR(p) process Y given the model data p the order, ($z_j,\ j = 1, \ldots, p$) the poles of the coefficient filter, σ^2 the error covariance, and M the number of observations. This derivation abreiviates that given in [5, Chapter 22]. The formulation is in terms of the autoregressive coefficients a, but in implementing this my input data is the zeros of the transfer function of this filter, which is $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}$. Notice that, a is completely determined by ($z_j,\ j = 1, \ldots, p$), by the identity

$$A(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) \cdots (1 - z_p z^{-1}), \qquad \forall z \in \mathbb{C}$$

so that each a_i is some symmetric polynomial of $z_1, \ldots, z_p$.

---

[2]The $\star$ denotes the usual convolution of infinite sequences

$$a \star Y = \left(\sum_{j=-\infty}^{\infty} a_j Y_{n-j},\ n > -\infty \right),$$

and for finite sequences "pad out" with infinite zeros.

[3]This is a consequence of a convolution theorem for the z-transform. The particular result used here is that for a WSS process X and an ℓ^1 filter $a = (\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots)$, with transfer function (z-transform) $A(z) = \sum_{n=-\infty}^{\infty} a_n z^{-n}$, then

$$S_{a \star X}(z) = A(z) S_X(z) A^*(z^{-*}).$$

The superscripts $-*$ together is an abbreviation for the complex conjugate of the multiplicative inverse, i.e.

$$z^{-*} = \frac{1}{z^*}.$$

A good source of this material is [1, Sec. 6.2]

To begin with, observe that since $\varepsilon \sim CN(0, \sigma^2 I_M)$ is Gaussian, $\tilde{Y} = (Y_n,\ n = 1, \ldots, M)$ is also Gaussian. So, let $\text{var}(\tilde{Y}) = \sigma^2 Q_M$. And the density of $\tilde{Y}$ my be written

$$f_{\tilde{Y}}(y) = (\pi\sigma_\varepsilon^2)^{-M} |Q_M|^{-1} \exp\left(\frac{-1}{\sigma_\varepsilon^2} y^* Q_M^{-1} y\right)$$

Here we think of $y = (y_n,\ n = 1, \ldots, M)$ as a sample or realization of the vector or finite time process $\tilde{Y}$. We seek to derive Q_M in terms of a and σ^2.

To that end, consider (1) specialized to the snapshot y, with corresponding error realization $\tilde{\varepsilon} = (\tilde{\varepsilon}_n,\ n = 1, 2, \ldots, M)$, written as

$$(a \star y)_n = \tilde{\varepsilon}_n \qquad \text{for } n = 1, \ldots, M$$

This can be written in matrix form as

$$A_* y^{(1)} + A y^{(2)} = \tilde{\varepsilon}^{(2)} \tag{2}$$

where $y^{(1)} = (y_1, \ldots, y_p)$ is the column vector of the first p observation of y and $y^{(2)} = (y_{p+1}, \ldots, y_M)$ is the column vector of the remaining $M - p$ observations of y. The column vectors $\tilde{\varepsilon}^{(1)}$ and $\tilde{\varepsilon}^{(2)}$ are analogously defined with respect to $\tilde{\varepsilon}$. The system (2) has only $M-p$ equations since the first p equations contain $\tilde{\varepsilon}^{(1)}$ which is not strictly speaking IID since the first p terms of y would only be partially convolved with a. The matrix A is the banded Toeplitz (lower triangular) $((M-p) \times (M-p))$-matrix whose first column is a followed by $M - (2p+1)$ zeros. And the matrix A_* is an extension of A due to the terms of $y^{(1)}$, which I know write out explicitly for clarity.

$$A_* = \begin{bmatrix} a_p & \ldots & a^{(1)} \\ \vdots & \ddots & \vdots \\ 0 & \ldots & a_p \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \\ 0 & \ldots & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ a_{p-1} & \ldots & 1 & 0 & \ldots & 0 & 0 \\ a_p & \ldots & a^{(1)} & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & a_p & a_{p-1} & \ldots & 1 & 0 \\ 0 & \ldots & 0 & a_p & \ldots & a^{(1)} & 1 \end{bmatrix}$$

Now rewriting (2), with a trivial addition, in block matrix form gives

$$\begin{bmatrix} y^{(1)} \\ \tilde{\varepsilon}^{(2)} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ A_* & A \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}, \tag{3}$$

which may be inverted as

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ -A^{-1}A_* & A^{-1} \end{bmatrix} \begin{bmatrix} y^{(1)} \\ \tilde{\varepsilon}^{(2)} \end{bmatrix}. \tag{4}$$

Now since

$$\text{var}\left(\begin{bmatrix} y^{(1)} \\ \tilde{\varepsilon}^{(2)} \end{bmatrix}\right) = \sigma^2 \begin{bmatrix} Q_p & 0 \\ 0 & I_{M-p} \end{bmatrix} \tag{5}$$

(Q_p to be determined later). We can write

$$\sigma^2 Q_M = \text{var}\left(\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}\right) = \sigma^2 \begin{bmatrix} I_p & 0 \\ -A^{-1}A_* & A^{-1} \end{bmatrix} \begin{bmatrix} Q_p & 0 \\ 0 & I_{M-p} \end{bmatrix} \begin{bmatrix} I_p & -A_*^* A^{-*} \\ 0 & A^{-*} \end{bmatrix}$$
$$= \sigma^2 \begin{bmatrix} Q_p & -Q_p A_*^* A^{-*} \\ -A^{-1}A_* Q_p & A^{-1}(A_* Q_p A_*^* + I_p)A^{-*} \end{bmatrix}.$$

Which can be inverted to

$$\sigma^{-2} Q_M^{-1} = \text{var}\left(\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}\right)^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} A_*^* A_* + Q_p^{-1} & A_*^* A \\ A^* A_* & A^* A \end{bmatrix}.$$

And thus we obtain

$$Q_M^{-1} = \begin{bmatrix} A_*^* A_* + Q_p^{-1} & A_*^* A \\ A^* A_* & A^* A \end{bmatrix}.$$

So that, we arrive at

$$y^* Q_M^{-1} y = y^{(1)*} Q_p^{-1} y^{(1)} + (A_* y^{(1)} + A y^{(2)})^* (A_* y^{(1)} + A y^{(2)}) \tag{6}$$

Observe that since y is assumed to be a realization of a segment of Y, and since Y is assumed to have started in the remote past, Q_M (the autocovariance of y)is Toeplitz and so Q_M^{-1} possesses symmetry about its NE-SW diagonal. Which means that

$$A_*^* A_* + Q_p^{-1} = A_{11} A_{11}^*$$

where A_{11} is the $p \times p$ principal submatrix of A and so,

$$Q_p^{-1} = A_{11} A_{11}^* - A_*^* A_*$$

Which allows us to write the likelihood of y is terms of only p, $(z_j,\ j = 1, 2, \ldots, p)$, σ^2, and M.

2.2. **Brief introduction to reversible jump Markov chain Monte Carlo.**

2.2.1. *Metropolis Hastings.* The Metropolis-Hastings (MH) algorithm allows us to sample from a target distribution π which we can evaluate (up to a multiplicative constant) but cannot sample from, such as in our case, a desired posterior distribution. The idea is to find a Markov chain which has as it's unique stationary distribution the target density. Recall that for a Markov chain over a statespace S, with transition density $p(x, y)$, a stationary distribution is one in which is on for which

$$\pi(y) = \int_S \pi(x) p(x, y) dx \quad \text{for all } y \in S. \tag{7}$$

Constructing a transition density p over $S \times S$ given only S and π is in general very difficult. But it turns out that the simplest A Markov chain is said to poses detailed balance if its stationary distribution and its transition density satisfies the following property for all $x, y \in S$

$$\pi(x) p(x, y) = \pi(y) p(y, x) \tag{8}$$

Detailed balance is a sufficient condition for a transition density p to satisfy (7). Observe

$$\int_S \pi(x)p(x,y)dx = \int_S \pi(y)p(y,x)dx = \pi(y)\int_S p(y,x)dx = \pi(y)$$

The MH algorithm takes in a proposal transition density q and adjusts it in a way that the adjustment is guaranteed to satisfy detailed balance for the target distribution. Thus the adjusted transition density describes a Markov chain whose stationary distribution is π.

I now describe the algorithm. Given the target distribution π which can be evaluated (up to a multiplicative constant) but which cannot be sampled from, the user provides a proposal transition density q which may be samples from at each point in the statespace. With this data together with an initial state $X_0 \in S$, the procedure precedes as in Algorithm 1.

Algorithm 1 Metropolis-Hastings

procedure MetropolisHastings(π,q,X_0)
  $X \leftarrow X_0$
  for $i = 1$ to N **do**
    Sample $Y \sim q(X_0, \cdot)$
    Evaluate acceptance probability α

$$\alpha(X,Y) = 1 \wedge \frac{\pi(Y)q(Y,X)}{\pi(X)q(X,Y)}$$

    Sample $u \sim \text{Unif}([0,1])$
    if $u \leq \alpha$ **then**
      Accept proposal: $X \leftarrow Y$

2.2.2. *Some issues with MH.* The above description of the algorithm is suitable for many purposes some exceptional cases however are served by the same concept but in a slightly different framework. Such is the problem posited in this paper. We are dealing with a posterior, our target distribution whose domain is subsets of many different dimensions. To elaborate, we desire the statespace to include $\mathcal{C}_1 = \{1\} \times \mathbb{D} \times \mathbb{R}_+$ on the one hand to including AR(1) processes but on the other hand AR(p) processes ought to be considered. These live in the parameter space $\mathcal{C}_2 = \{2\} \times \mathbb{D}^2 \times \mathbb{R}_+$. So that if I wanted to transition from $x \in \mathcal{C}_1$ to $x' \in \mathcal{C}_2$, I met with a complexity following the original Metropolis Hastings formulation, in that it asks me to compute

$$\frac{\pi(x)q(x',x)}{\pi(x')q(x,x')}.$$

This draws my attention to the fact that q is a function on many different spaces and therefore must be defined with that in mind, and the acceptance ratio α must in turn be defined I a way that produces detailed balance.

Another problem is that above we integrate using densities times the Lebesgue measure over the natural ambient Euclidean space, but what if the set spans spaces of different dimension the measure we would like to use would have to be in some way extended.

Principally, the goal of reversible jump Markov chain Monte Carlo (RJMCMC) is to allow for the Markov chain to have a statespace that includes spaces of different dimension. Such

as

$$\mathcal{C} = \bigcup_{k=1}^{K} \mathcal{C}_k \qquad \text{where } \mathcal{C}_k \subset \mathbb{R}^{n_k},\ K \leq \infty$$

One way to approach these subtleties is to consider extending detailed balance in an integral sense. Given here,

$$\int_A \int_{A'} \pi(x)p(x,x')dx'dx = \int_A \int_{A'} \pi(x')p(x',x)dx'dx \qquad \text{for all } A \in \mathcal{C}_k,\ A' \in \mathcal{C}_{k'} \tag{9}$$

Notice that if (8) holds then (9) will hold. Conversely, if (9) holds for all admissible k, k', then (8) will hold[4].

This will provide a new way of seeing the old HM algorithm. Consider sets $A \subset \mathcal{C}_k \subset \mathbb{R}^{n_k}$ and $A' \subset \mathcal{C}_{k'} \subset \mathbb{R}^{n'_k}$. If $n_k = n'_k$ the acceptance ratio can be computed the same as regular MH, so suppose then, with no loss of generality $n_k < n'_k = r + n_k$ And consider now this extended sort of detailed balance. Now, define a transition kernel $P : \mathbb{C}_k \times B(\mathbb{C}'_k) \to \mathbb{R}_+$ by

$$P(x, A') = \int_{A'} p(x,x')dx'$$

this gives the density with respect to x over $\mathcal{C}_k$ that the probability that the adjusted proposal Markov chain transitions from a point x to a set A'. Using the basic MH framework we see that this would be written as follows

$$P(x, A') = \int_{A'} q(x,x')\alpha(x,x')dx' + I_{x \in A'} \cdot \left[\int q(x,x')[1 - \alpha(x,x')]dx' \right]. \tag{10}$$

If we substitute $P(x, A')$ for the inner iterated integral we get on the left-hand side of (9) which becomes

$$\begin{aligned} &\int_A \pi(x) \int_{A'} q(x,x')\alpha(x,x')dx'dx + \int_A \pi(x) I_{x \in A'} \int q(x,x')[1 - \alpha(x,x')]dx'dx \\ &\qquad = \int_A \int_{A'} \pi(x)q(x,x')\alpha(x,x')dx'dx + 0. \end{aligned}$$

Note that since A and A' live in different dimensions their intersection is empty. And, repeatimg the procedure for the right hand side of (9) gives,

$$\begin{aligned} &\int_{A'} \pi(x') \int_A q(x',x)\alpha(x',x)dxdx' + \int_{A'} \pi(x') I_{x' \in A} \int q(x',x)[1 - \alpha(x',x)]dxdx' \\ &\qquad = \int_A \int_{A'} \pi(x')q(x',x)\alpha(x',x)dx'dx + 0. \end{aligned}$$

[4]To see this, supposing p and π to be Borel measurable, let $E = \{(x,y) \in \mathcal{C}_k \times \mathcal{C}_{k'} | \pi(x)p(x,y) \leq \pi(y)p(y,x)\}$. And let R_i be a countable collection of closed, non-overlapping rectangles such that $R_i = A \times A'$ for some $A \in \mathcal{C}_k$ and some $A' \in \mathcal{C}_{k'}$ and $E = \bigcup_{i=1}^{\infty} R_i$ then note that

$$\begin{aligned} \iint_E |\pi(x)p(x,y) - \pi(y)p(y,x)|\ dxdy &= \iint_E \pi(x)p(x,y) - \pi(y)p(y,x)\ dxdy \\ &= \sum_{i=1}^{\infty} \iint_{R_i} \pi(x)p(x,y) - \pi(y)p(y,x)\ dxdy = 0 \end{aligned}$$

So, that $\pi(x)p(x,y) = \pi(y)p(y,x)$ a.e. on E. This can be shown for E^c as well.

So, for integral detailed balance to hold it is sufficient for

$$\int_A \int_{A'} \pi(x)q(x,x')\alpha(x,x')dx'dx = \int_A \int_{A'} \pi(x')q(x',x)\alpha(x',x)dx'dx. \tag{11}$$

It is usually possible to sample r random numbers u from some known (joint) distribution g and then propose the new state x' by a suitable deterministic function h, such that $(x',u') = h(x,u)$. The variable u' is r' random numbers generated from a know distribution g' so that the reverse move from x' to x uses u' to recover u, by h' the inverse of h. This, of course, requires $n_k + r = n_{k'} + r'$ which is referred to as the dimension matching condition. Observe that in conjunction with x, u determines x', so that x and u parameterize x' and we can therefore integrate solely with respect to x and u on the left-hand side, and similarly for the right-hand side. In this way the the proposal density $q(x,x')$ can be identified with $g(u)$ and $q(x',x)$ with $g'(u')$ so that (11) becomes

$$\int_{x\in A} \int_{x\in A'} \pi(x)g(u)\alpha(x,x')dxdu = \int_{x\in A} \int_{x\in A'} \pi(x')g'(u')\alpha(x',x)dx'du'. \tag{12}$$

If h and h' is differentiable (h is diffeomorphic), then a standard change-of-variables can be applied to put both $(n+r)$-dimensional integrals in the same variable space and find that equality holds if

$$\pi(x)g(u)\alpha(x,x') = \pi(x')g'(u')\alpha(x',x)\left|\frac{\partial(x',u')}{\partial(x,u)}\right|.$$

And so, we can pick α to be

$$\alpha(x,x') = 1 \wedge \frac{\pi(x')g'(u')}{\pi(x)g(u)}\left|\frac{\partial(x',u')}{\partial(x,u)}\right|. \tag{13}$$

2.3. **Bayesian inference and MCMC.** Bayesian inference can be used in conjunction with MCMC in model determination. Suppose we have data y and a model with parameters θ. We would like to understand how the parameters are distributed given a realization of the "model", the assumption being y is drawn for our proposed model. Bayes rule says

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{14}$$

In (14) the left-hand side is called the posterior, and the right-hand side comprises, from right to left, the likelihood and the priors. Sometimes, we is the case above, we have a formula for the likelihood and some principles way of assigning a prior. In that case one can study the posterior this way. We use Markov chain Monte Carlo to same form distributions otherwise may not be able to sample from. And so using that technique together with knowledge of the likelihood and prior we may be able to create a markov chain to sample from the posterior.

In the present context, we are interested in

$$p(p, \mathbf{z}, \sigma^2|y)$$

. This is the probability density over the parameter space conditioned on the observed data. The parameters here are random variables, with their own distributions, the priors, but by virtue of the assumption that y was drawn form an $AR(p)$ model with these parameter values drawn for these distributions there is dependence between y and the parameters $p, \mathbf{z}, \sigma^2$. We

exploit that dependence in conditioning on y and getting the conditional distribution of the parameters. Basyes rule allows us to re-frame this object to a more approachable one. We write

$$p(p, \mathbf{z}, \sigma^2|y) \propto p(y|p, \mathbf{z}, \sigma^2)p(p, \mathbf{z}, \sigma^2) \tag{15}$$

$$= p(y|p, \mathbf{z}, \sigma^2)p(p, \mathbf{z})p(\sigma^2) \tag{16}$$

$$\propto p(y|p, \mathbf{z}, \sigma^2)p(\mathbf{z}|p)p(p)p(\sigma^2). \tag{17}$$

In our case it will be reasonable to assume that error variance is independent with the order of the model and the poles of the model. To apply this to the acceptance probability from (13) we get

$$\begin{aligned}
\alpha &= 1 \wedge \frac{\pi(p', v', z'|y)}{\pi(p, v, z|y)} \cdot \frac{q(v, z|v', z', p, p')q(p|p')}{q(v', z'|v, z, p, p')q(p'|p)} \left|\frac{\partial(v', z')}{\partial(v', z')}\right| \\
&= 1 \wedge \frac{\pi(y|v', z', p')\pi(v', z'|p')\pi(p')}{\pi(y|v, z, p)\pi(v, z|p)\pi(p)} \cdot \frac{q(v, z|v', z', p, p')q(p|p')}{q(v', z'|v, z, p, p')q(p'|p)} \left|\frac{\partial(v', z')}{\partial(v', z')}\right| \\
&= 1 \wedge \frac{\pi(y|v', z', p')}{\pi(y|v, z, p)} \cdot \frac{\pi(v', z'|p')\pi(p')}{\pi(v, z|p)\pi(p)} \cdot \frac{q(v, z|v', z', p, p')q(p|p')}{q(v', z'|v, z, p, p')q(p'|p)} \left|\frac{\partial(v', z')}{\partial(v', z')}\right|
\end{aligned}$$

$$\alpha = 1 \wedge (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian})$$

3. Specification of the algorithm

In this section I implement formulate the Markov chain and implement it. I begin by with an overview clarifying what the statespace is and how, in general the process will move around it. I then go into details starting with the priors of the parameters of the statespace. Then the moves types are identified and the probability with which they are chosen are set forth. Lastly I take each move type in turn and, along with move specific details, I provide the proposal density and acceptance probabilities. In the next section he chain is then implemented and results are reported.

The goal is to produce a Markov chain that tends toward the stationary distribution of which is the posterior $p(p, \mathbf{z}, \sigma^2|y)$ (from (15)) which means the chain spends most of its time in areas correspond to parameters of high probability given the time series realization. Therefore, the statespace will be

$$\mathcal{C} = \bigcup_{p=1}^{p_{\max}} \mathcal{C}_p \qquad \text{where } \mathcal{C}_p = \{p\} \times \mathbb{R}^+ \times \mathbb{D}^p$$

$p_{\max}$ is a hyper parameter to be determined by the user. A generic element of this space will be denoted

$$x = (p, v, z_1, z_2, \ldots, z_p) \in \mathcal{C}.$$

Here p is the order of the purposed AR(p) model, v ($= \sigma^2$) is the variance of the white noise error process, and $\mathbf{z} = (z_1, z_2, \ldots, z_p)$ is the location of the poles of the model.

The Markov chain will be denoted $X_n = (P_n, V_n, \mathbf{Z}_n)$ where $\mathbf{Z}_n$ are random vectors of varying length.

3.1. **Priors.** The order P will be a Poisson distributed with rate λ, conditioned on $P \leq p_{\max}$ so that

$$f_P(p) = \frac{\lambda^p e^{-\lambda}}{p!} \cdot I_{p\in[0,p_{\max}]} \left(\sum_{j=0}^{p_{\max}} \frac{\lambda^j e^{-\lambda}}{j!} \right)^{-1}$$
$$\propto \frac{\lambda^p e^{-\lambda}}{p!} \cdot I_{p\in[0,p_{\max}]}$$

For the variance V, let $u \sim \text{Unif}([-\beta, \beta])$, $\beta > 0$ and let

$$V = e^u \qquad \text{so that} \qquad f_V(v) = \frac{1}{2\beta v} \mathbf{1}_{[-\beta,\beta]}(\log(v)).$$

The location of the poles will be independently and uniformly distributed on the unit disk. So, for a given order p the prior of the pole locations $\mathbf{Z} = (Z_1, Z_2, \dots Z_p)$ can be written as

$$f_P(z_1, z_2, \dots, z_p) = \frac{1}{\pi^p} \prod_{j=1}^{p} \mathbf{1}_{\mathbb{D}}(z_j)$$

since the p poles are independent,

$$f_{Z_j}(z_j) = \frac{1}{\pi} \mathbf{1}_{\mathbb{D}}(z_j) \qquad \text{for } j = 1, \dots, p.$$

3.2. **Moves and move types.** There are three move types, (1) change in variance, (2) change in pole position, and (3) birth or death of pole. The change of variance occurs with probability ν_p (p is the number of poles). Change in pole position occurs with probability π_p and birth or death of pole occurs with probabilities b_p, and d_p, respectively. More precisely, a move from $\mathcal{C}_p$ to $\mathcal{C}_{p+1}$ occurs with probability b_p and from $\mathcal{C}_{p+1}$ to $\mathcal{C}_p$ occurs with probability d_p The probabilities observe the following conditions

(a) $\beta_p + \pi_p + b_p + d_p = 1$ for all p
(b) $d_0 = \pi_0 = b_{p_{\max}} = 0$
(c) $b_p = c \min\{1, f_P(p+1)/f_P(p)\}$, and $d_p = c \min\{1, f_P(p)/f_P(p+1)\}$

for some c as large as possible so that $b_p + d_p \leq 0.9$

3.2.1. *Moves and their Acceptance probabilities.* I will now discuss the moves themselves.

A change in variance is a move within $\mathcal{C}_p$. The new variance V' will be so that

$$\log(V'/V) \sim \text{Unif}([-\beta, \beta])$$

meaning $V' = Ve^u$ where $u \sim \text{Unif}([-\beta, \beta])$ and so,

$$f_{V'|V}(v', v) = \begin{cases} \frac{1}{2\beta v'}, & v' \in \left[ve^{-\beta}, ve^{\beta}\right] \\ 0, & \text{otherwise} \end{cases}$$

Observe that if

$$v' \in \left[ve^{-\beta}, ve^{\beta}\right] \qquad \text{then} \qquad v \in \left[v'e^{-\beta}, v'e^{\beta}\right]$$

And since V' is contrived with probability one to fall in $\left[Ve^{-\beta}, Ve^{\beta}\right]$ the proposal ratio is

$$\frac{f_{V|V'}(v,v')}{f_{V'|V}(v',v)} = \frac{\frac{1}{2\beta v}}{\frac{1}{2\beta v'}} = \frac{v'}{v}$$

The likelihood ratio is

$$\frac{(\pi v')^{-M}|Q_M|^{-1}\exp\left(\frac{-1}{v'}y^*Q_m^{-1}y\right)}{(\pi v)^{-M}|Q_M|^{-1}\exp\left(\frac{-1}{v}y^*Q_m^{-1}y\right)} = \left(\frac{v}{v'}\right)^M \exp\left(-y^*Q_m^{-1}y(1/v - 1/v')\right) \tag{18}$$

The acceptance ratio is computed to be

$$\begin{aligned}
\alpha &= 1 \wedge (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian}) \\
&= 1 \wedge \left(\frac{v}{v'}\right)^M \exp\left(-y^*Q_m^{-1}y(1/v - 1/v')\right) \times \frac{\frac{1}{2\beta v'}\mathbf{1}_{[-\beta,\beta]}(\log(v'))}{\frac{1}{2\beta v}\mathbf{1}_{[-\beta,\beta]}(\log(v))} \times \frac{v'}{v} \times 1 \\
&= 1 \wedge \left(\frac{v}{v'}\right)^M \exp\left(-y^*Q_m^{-1}y(1/v - 1/v')\right)
\end{aligned}$$

For a change in position of a pole randomly select $j = 1, \ldots, p$ (uniformly) and the pole Z_j will be perturbed by $\tilde{u} \sim CN(Z_j, \hat{\pi})$ conditioned on $Z_j' = Z_j + \tilde{u} \in \mathbb{D}$. This gives

$$f_{Z_j'|Z_j}(z', z) \propto \begin{cases} \frac{1}{\pi\hat{\pi}}\exp\left(\frac{-1}{\hat{\pi}}|z'-z|^2\right)[I(z)]^{-1}, & z' \in \mathbb{D} \\ 0, & \text{otherwise} \end{cases}$$

where $I(z) = \dfrac{1}{\pi\hat{\pi}}\displaystyle\int_{\mathbb{D}} \exp\left(\frac{-1}{\hat{\pi}}|z'-z|^2\right) dz'$ w.r.t Lebesgue measure on $\mathbb{C}$. A calculation shows that

$$I(z) = \int_0^{2\pi} 1 - \exp\left(\frac{-1}{\hat{\pi}}(1 + |z|^2 - 2|z|\cos t)\right) dt.$$

The prior ratio is unity since practically z' is enforced to be in the unit disk. The proposal ratio is given by $I(z')/I(z)$ and the likelihood ration is

$$\frac{|Q_M'|^{-1}\exp\left(\frac{-1}{v'}y^*(Q_m')^{-1}y\right)}{|Q_M|^{-1}\exp\left(\frac{-1}{v}y^*Q_m^{-1}y\right)} \tag{19}$$

where Q_M' is computed with z_j' in place of z_j. And so, the acceptance ratio reduces to

$$\alpha = 1 \wedge \frac{|Q_M'|^{-1}\exp\left(\frac{-1}{v'}y^*(Q_m')^{-1}y\right)}{|Q_M|^{-1}\exp\left(\frac{-1}{v}y^*Q_m^{-1}y\right)} \times \frac{I(z')}{I(z)}$$

The the birth and death of a pole we need to consider the move together since one reverses the other and vice versa. For the birth a pole we move from $\mathcal{C}_p$ to $\mathcal{C}_{p+1}$ by appending a pole drawn from the uniform distribution on the unit disk. We are changing dimensions so we need to check that we meet the dimension matching condition. To check this we need to describe the reverse move, the death of a pole. The death of a pole is the deletion of a pole chosen at random. This is a very simple case, in the context of Section 2.2 we simple set $u \sim Unif(\mathbb{D})$ to be the new pole, then the map to new state (from $\mathcal{C}_p$) to $\mathcal{C}_{p+1}$ is linear and almost the identity (were it not for the value of p changing to $p+1$). The inverse of this

map (from $\mathcal{C}_{p+1}$ to $\mathcal{C}_p$) simply maps z_j to u' where it gets discarded. The Jacobian therefore is 1.

For the priors ratio, the priors to consider can be seen under (17), we need not consider $p(\sigma^2)$ since v remains unchanged by this move and will not affect the ratio. However, we have

$$p(\mathbf{z}'|p') = \frac{1}{\pi^{p+1}} \qquad \text{and} \qquad p(\mathbf{z}|p) = \frac{1}{\pi^p}$$

and

$$\frac{p(p')}{p(p)} = \frac{\lambda^{p+1}e^{-\lambda}}{(p+1)p!} \cdot \frac{p!}{\lambda^p e^{-\lambda}} = \frac{\lambda}{p+1}$$

and the proposal ratio is

$$\frac{q((p,\mathbf{z}),(p',\mathbf{z}'))}{q((p',\mathbf{z}'),(p,\mathbf{z}))} = \frac{d_p}{b_p/\pi}$$

Putting all this together gets an acceptance probability for the birth of a pole of

$$\alpha = 1 \wedge \frac{|Q'_M|^{-1}\exp\left(\frac{-1}{v'}y^*(Q'_m)^{-1}y\right)}{|Q_M|^{-1}\exp\left(\frac{-1}{v}y^*Q_m^{-1}y\right)} \times \frac{\lambda}{p+1}\frac{d_p}{b_p} \tag{20}$$

Similarly for the death of a pole,

$$\alpha = 1 \wedge \frac{|Q'_M|^{-1}\exp\left(\frac{-1}{v'}y^*(Q'_m)^{-1}y\right)}{|Q_M|^{-1}\exp\left(\frac{-1}{v}y^*Q_m^{-1}y\right)} \times \frac{p+1}{\lambda}\frac{b_p}{d_p} \tag{21}$$

4. Experiments and Results

This work is unfinished, however, I here describe my progress in implementing this scheme. I began with a known realization of an $AR(1)$ process with a pole at $z_1 = -0.5$ and whit noise variance set to $\sigma^2 = 1$.

I set the parameters to be

$$\lambda = 1 \qquad \beta = \frac{1}{5} \qquad \hat{\pi} = 0.2$$

For this first run p was fixed, through out, $b_k = d_k = 0$ for all k. I also $\beta_{p=1} = 0.2$,

Figure 1 shows 2 independent Markov chains. And exhibits two major problems. First is that no matter the starting condition the chain tends towards the point 1 on the real axis, seeming to usually reject moves not in that direction. Second is that the variance tends towards zero, exponentially.

5. Discussion

With more time I hope to identify the causes of the aforementioned unexpected behavior.

References

[1] Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*. Number BOOK. Prentice Hall, 2000.

[2] Ali H Sayed and Thomas Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.

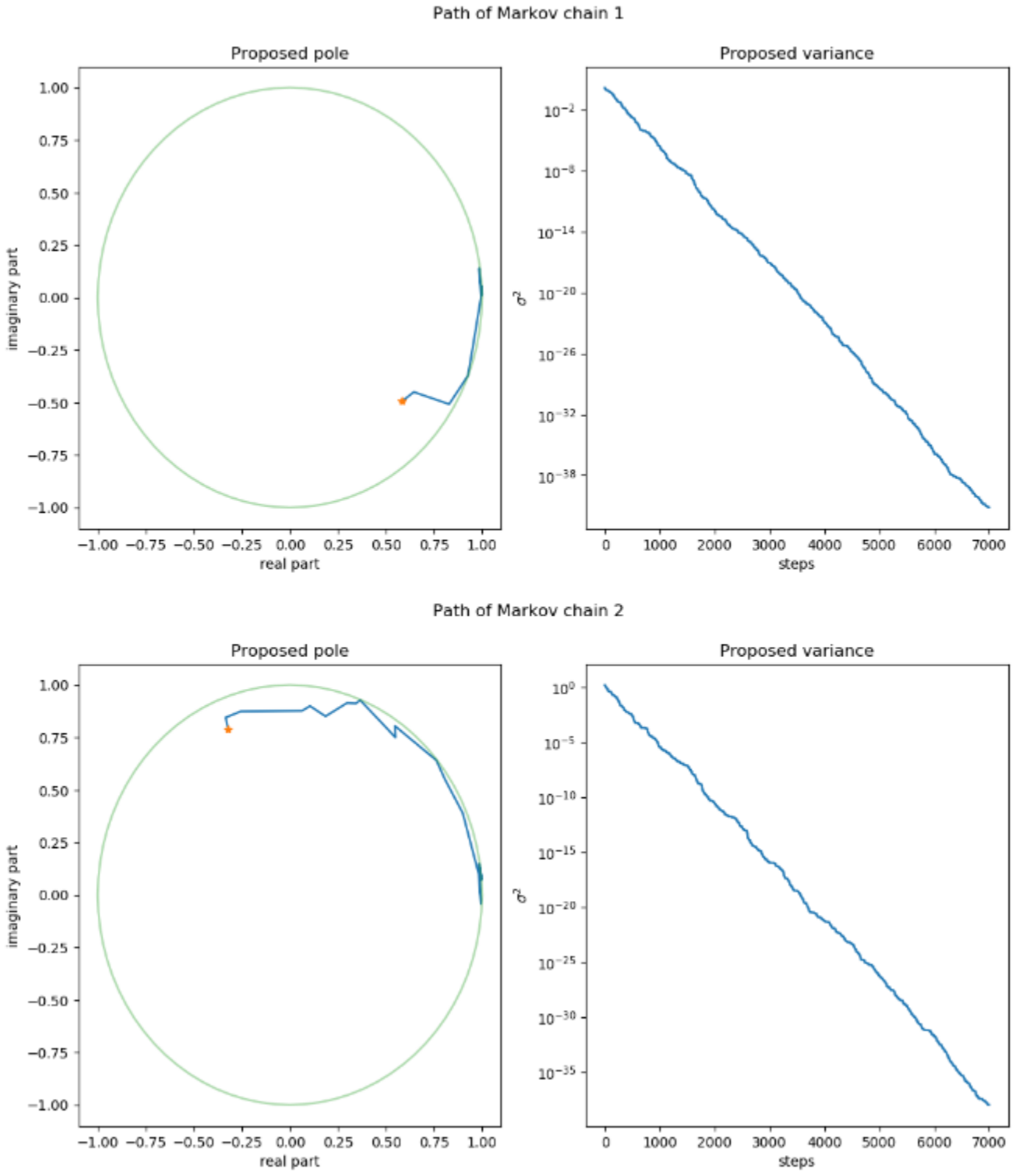

FIGURE 1. Two realizations of the Markov chain with fixed number of poles.

[3] Gigla Janashia, Edem Lagvilava, and Lasha Ephremidze. A new method of matrix spectral factorization. *IEEE Transactions on information theory*, 57(4):2318–2326, 2011.

[4] Yuzo Hosoya and Taro Takimoto. A numerical method for factorizing the rational spectral density matrix. *Journal of Time Series Analysis*, 31(4):229–240, 2010.

[5] David Stephen Geoffrey Pollock, Richard C Green, and Truong Nguyen. *Handbook of time series analysis, signal processing, and dynamics*. Elsevier, 1999.

[6] Alessandro Martini, Stefano Schmidt, and Walter Del Pozzo. Maximum entropy spectral analysis: a case study, 2021.

[7] Peter J Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[8] David I Hastie and Peter J Green. Model choice using reversible jump markov chain monte carlo. *Statistica Neerlandica*, 66(3):309–338, 2012.

[9] Jem Corcoran. Reversible jump mcmc (rjmcmc), Nov 2018. unpublished.

[10] Jianqing Fan and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media, 2008.