12-2022

Lindley Processes with Correlated Changes

John Grant
jwgrant@clemson.edu

Lindley Processes with Correlated Changes

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

---

by
John William Grant
December 2022

Accepted by:
Dr. Xin Liu, Committee Chair
Dr. Robert Lund
Dr. Peter Kiessler
Dr. Brian Fralix

Abstract

This dissertation studies a Lindley random walk model when the increment process driving the walk is strictly stationary. Lindley random walks govern customer waiting times in many queueing models and several natural and business processes, including snow depths, frozen soil depths, inventory quantities, etc. Probabilistic properties of a Lindley process with time-correlated stationary changes are explored. We provide a streamlined argument that the process admits a limiting stationary distribution when the mean of the incremental changes is negative and that the Lindley process is strictly stationary when starting from this stationary distribution. The Markov characteristics of the process are explored when the change process has a Markov structure of first or higher order. A derivation of the model's likelihood is given when the change process obeys a pth order autoregression. Due to the unwieldy nature of this likelihood, a particle filtering method of evaluating and optimizing it is devised and studied via simulation.

Dedication

This dissertation is dedicated to my grandmother, Anna Y. Cochran, who helped to instill in me a lifelong love of learning.

Acknowledgments

This work was conducted under the direction of Dr. Robert Lund, Dr. Xin Liu, and Dr. Jonathan Woody, with some collaboration with Jiajie Kong at U.C. Santa Cruz. For general support and inspiration, I would like to thank my parents Ron and Kathy Grant, my grandparents Robert and Ann Cochran and Ralph and Mae Grant, my brother Jeff Grant, Rich Lotstein, Susan Hassol, Jonathan Woody, and Clifton Avery. I would like to express my deepest appreciation and love to my fiancee, Alina Lotstein.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter, we will provide the foundation upon which our later work will depend. We will explore the background of an important class of time series models, introduce maximum likelihood estimation, and then provide a brief overview of Lindley processes.

1.1 Overview of Stationary ARMA Processes

In this section we give an overview of one of the most important classes of time series models, the stationary autoregressive moving average or ARMA processes. The development given here will closely follow the presentation by Brockwell and Davis in *Introduction to Time Series and Forecasting* and *Time Series: Theory and Methods* [3, 4]. The structure and properties of this family of processes make it one of the most useful models for time-series data. If we want a model that is useful for making predictions, then we seek a model with some properties that remain constant over time. When analyzing time series, we are in general considering a series that contains a random component. If certain key statistical properties are stationary or constant over time, then we can establish techniques for forecasting future values. In order to develop the concept of stationarity, we first introduce the autocovariance and autocorrelation functions.

When considering finite collections of random variables, the covariance matrix is often useful for examining the dependence between pairs of variables. The autocovariance function allows us to extend the concept of the covariance matrix to an infinite time series. The autocovariance function or ACVF for a time series $\{X_t\}$ is a function that gives the covariance of the series with itself at

pairs of time points. Let $\{X_t, 0, \pm 1, \dots\}$ be a time series with $E[X_t^2] < \infty$. The mean function of $\{X_t\}$ is defined as

$$\mu_X(t) = E[X_t].$$

The autocovariance function of $\{X_t\}$ is defined as

$$\gamma(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))].$$

Speaking informally, a time series $\{X_t\}$ is considered stationary if it has similar statistical properties to the series $\{X_{t+h}\}$ where h is an integer. In other words, the series $\{X_{t+h}\}$ has been shifted by h time steps. We will be concerned with two main types of stationarity, strict stationarity and weak or second-order stationarity. A time series $\{X_t\}$ is said to be strictly stationary if, for all h, $(X_1, \dots, X_n)$ has the same joint distribution as $(X_{1+h}, \dots, X_{n+h})$. This type of stationarity will be significant in the arguments we employ in subsequent sections of this dissertation. In the case of weak stationarity, we are concerned only with properties that depend on the first and second-order moments of $\{X_t\}$. A time series $\{X_t\}$ where $E[X_t^2] < \infty$ for all t is said to be weakly stationary if it meets the following conditions:

$$\text{(i) } \mu_X(t) = E[X_t] \text{ is independent of } t,$$

and

$$\text{(ii) } \gamma_X(t + h, t) \text{ is independent of } t \text{ for each } h.$$

One can easily show that if $\{X_t\}$ is strictly stationary and $E[X_t^2] < \infty$ for all t, then $\{X_t\}$ must also be weakly stationary. However, weak stationarity of a process does not necessarily imply strict stationarity. In general, when a time series is referred to as stationary it is understood to mean weakly stationary unless otherwise specified. As a direct consequence of the second weak stationarity condition given above, we may write the autocovariance function of a stationary time series $\{X_t\}$ as a function of one variable:

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t + h, t).$$

$\gamma_X(h)$ is referred to as the value for the autocovariance function of $\{X_t\}$ at lag h.

Another function which is closely related to the autocovariance function is the autocorrelation function. It gives the Pearson correlation between values of the process at different time points. If $\{X_t\}$ is a stationary time series with autocovariance function $\gamma_X(h)$ the autocorrelation function or ACF is given by

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Cor}(X_{t+h}, X_t). \tag{1.1}$$

The following example will illustrate the concepts of ACVF and ACF, as well as introduce a process that will be important as we examine ARMA models in general.

Example 1.1.1. White Noise Process

A time series $\{X_t\}$ is a white noise process if it is a series of uncorrelated random variables such that $E[X_t^2] = \sigma^2 < \infty$ and $E[X_t] = 0$ for all t. One can easily see that this process fits the stationarity conditions given above. Condition (i) is obviously satisfied since $E[X_t] = 0$ for all t. Since we have assumed that $\{X_t\}$ is uncorrelated then its autocovariance function is given by

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2, & \text{if } h = 0 \\ 0, & \text{if } h \neq 0 \end{cases}$$

which is independent of t, and so condition (ii) is satisfied. Thus $\{X_t\}$ is (weakly) stationary. We indicate a white noise process by the notation

$$\{X_t\} \sim \text{WN}(0, \sigma^2).$$

A special type of white noise process is called iid noise. Here the assumptions are the same as those for the white noise process, but we specify that $\{X_t\}$ is an iid (independent and identically distributed) sequence. We indicate such a process by

$$\{X_t\} \sim \text{IID}(0, \sigma^2).$$

It is clear that every iid noise process is also a white noise process. However, the converse is not true, as lack of correlation does not necessarily imply independence.

Example 1.1.2. A Random Walk

Let $S_0 = 0$ and suppose that

$$S_t = X_1 + X_2 + \cdots + X_t, \quad \text{for } t = 1, 2, \ldots,$$

where $\{X_t\} \sim \text{IID}(0, \sigma^2)$. Clearly $E[S_t] = 0$ for all t. Also note that

$$\begin{aligned} E[S_t^2] &= \text{Var}(S_t) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_t) \\ &= t\sigma^2. \end{aligned}$$

To determine if the process $\{S_t\}$ is stationary, we need to check that the autocovariance function satisfies condition (ii) above. For $h \geq 0$, the ACVF is given by

$$\begin{aligned} \gamma_S(t+h, t) &= \text{Cov}(S_{t+h}, S_t) \\ &= \text{Cov}(S_t + X_{t+1} + \cdots + X_{t+h}, S_t) \\ &= \text{Cov}(S_t, S_t) \\ &= t\sigma^2 \end{aligned}$$

which is not independent of t, so we conclude that $\{S_t\}$ is not stationary.

To further illustrate the autocovariance function and the concept of stationarity, we now introduce two simple examples of the more generalized autoregressive moving-average or ARMA process that we will later examine in greater detail.

Example 1.1.3. MA(1) Process
Let $\{X_t\}$ be defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \ldots,$$

where $Z_t \sim \text{WN}(0, \sigma^2)$ and θ is a constant real number. We refer to such a series as a first-order moving average or MA(1) process. It is clear from the defintion that $E[X_t] = 0$ for all t. By making use of the linearity property of covariances, we find that $E[X_t^2] = Var(X_t) = \sigma^2(1 + \theta^2) < \infty$. The

autocovariance function is given by

$$\gamma_X(t+h,t) = \begin{cases} \sigma^2(1+\theta^2), & \text{if } h = 0 \\ \sigma^2\theta, & \text{if } h = \pm 1 \\ 0, & \text{if } |h| > 1. \end{cases}$$

Observing that $\{X_t\}$ meets all the proper requirements, we can conclude that it is a stationary process. Using the definition of ACF in (1.1), we find that the autocorrelation function is

$$\gamma_X(t+h,t) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{\theta}{(1+\theta^2)}, & \text{if } h = \pm 1 \\ 0, & \text{if } |h| > 1. \end{cases}$$

In our next example, we introduce the AR(1) process. As mentioned above, this process is a special case of the ARMA(p,q) process, and is the most simple example of an AR(p) process. The AR(p) process will figure prominently in the theory we will develop later in this dissertation.

Example 1.1.4. AR(1) Process

Suppose that $\{X_t\}$ is a time series satisfying the equation

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \ldots, \tag{1.2}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$. For now, we will assume that $\{X_t\}$ is a stationary series. Such a process is known as a first-order autoregression or AR(1) process. We will show later that under these conditions a unique solution exits for (1.2). First note that since $E[Z_t] = 0$, if we take expectations on both sides of (1.2), then it follows immediately that $E[X_t] = 0$. Next we find the autocorrelation function for $\{X_t\}$ by multiplying both sides of (1.2) by X_{t-h} when $h > 0$ to obtain

$$\begin{aligned} \gamma_X(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov}(\phi X_{t-1}, X_{t-h}) + \text{Cov}(Z_t, X_{t-h}) \\ &= \phi\gamma_X(h-1) + 0 = \cdots = \phi^h \gamma_X(0) \end{aligned}$$

where in the last step we have recursed $h-1$ times. Now note that

$$\begin{aligned}\gamma_X(h) &= \text{Cov}(X_{t+h}, X_t) \\ &= \text{Cov}(X_t, X_{t+h}) \\ &= \gamma_X(-h)\end{aligned}$$

and using the definition of ACF in (1.1) we find that

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h = 0, \pm 1, \ldots \tag{1.3}$$

If we want to find an expression for the ACVF at lag h, $\gamma_X(h)$, we will need to compute $\gamma_X(0)$. We can do this by using the linearity of the covariance function and noting that Z_t is uncorrelated with X_{t-1} by assumption. Thus

$$\begin{aligned}\gamma_X(0) &= \text{Cov}(X_t, X_t) \\ &= \text{Cov}(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) \\ &= \phi^2 \gamma_X(0) + \sigma^2 \\ \implies \gamma_X(0) &= \frac{\sigma^2}{(1-\phi^2)}.\end{aligned}$$

We can then substitute this expression into (1.3) to obtain the ACVF.

Although the topics treated in this dissertation are largely theoretical, we will touch briefly on some practical considerations that are important when finding a time series model for observed data. In the previous examples we were able to find the ACVF and ACF for several simple examples of time series. However, when modeling actual data, we will start with a series of observations. It is often desirable to determine if the data are realizations of a stationary time series model. In order to select an appropriate model, we would like to examine the degree of dependence in the data at different time lags. To this end, the sample autocovariance and autocorrelation functions (sample ACVF and sample ACF) are among the most important tools. We may consider the definitions of these functions of the data as sample analogues of the definitions given above for the ACVF and ACF of a stationary time series.

Given the data $x_1, x_2, \ldots, x_n$, we start with the usual definition of the **sample mean**, $\bar{x}$

$$\bar{x} := \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is defined as

$$\hat{\gamma}(h) := n^{-1}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n$$

By analogy with the definition of autocorrelation function, we define the **sample autocorrelation function** as

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad -n < h < n$$

These functions have been carefully defined so that the sample covariance matrix $\hat{\Gamma}(h)_n := [\hat{\gamma}(i-j)]_{i,j=1}^{n}$, and sample correlation matrix $\hat{R}_n := [\hat{\rho}(i-j)]_{i,j=1}^{n}$, have the desirable property of being nonnegative definite. A thorough treatment of the practical applications of these functions would carry us beyond the scope of this work, but they are mentioned here for completeness.

In order to develop generalized methods for studying stationary processes, we now introduce the class of time series models known as linear processes. It can be shown by an important result known as Wold's decomposition that every weakly stationary process is either a linear process or can be decomposed into the sum of a linear process and a deterministic component. The autoregressive moving-average or ARMA models that we will focus on are an important subclass of linear models.

A time series $\{X_t\}$ is a **linear process** if it can be written in the form

$$X_t = \sum_{j=-\infty}^{\infty} \beta_j Z_{t-j} \tag{1.4}$$

for all t, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $\{\beta_j\}$ is a sequence of constants such that

$$\sum_{j=-\infty}^{\infty} |\beta_j| < \infty. \tag{1.5}$$

For compactness of notation, we introduce the backward shift operator B, which operates on any

element of a time series to give the previous element:

$$BX_t = X_{t-1}.$$

Raising the backward shift operator to an arbitrary integer power k gives

$$B^k X_t = X_{t-k}.$$

In terms of the backward shift operator, (1.4) can be written as

$$X_t = \beta(B)Z_t \tag{1.6}$$

where $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$. The operator $\beta(B)$ is sometimes referred to as a linear filter in the sense that if we apply the operator to the input white noise series $\{Z_t\}$ we get the output series $\{X_t\}$.

The condition in (1.5) is important for several reasons. It guarantees that the infinite sum in (1.4) converges with probability one. To see this, note that since $E|Z_t| < \sigma$ we have

$$\begin{aligned} E|X_t| &\leq \sum_{j=-\infty}^{\infty} (|\beta_t||E|Z_{t-j}|) \\ &\leq \left(\sum_{j=-\infty}^{\infty} |\beta_j| \right) \sigma \\ &< \infty. \end{aligned}$$

One can also show that condition (1.5) ensures $\sum_{j=-\infty}^{\infty} \beta_j^2 < \infty$. This in turn guarantees that the series (1.4) converges in mean square, that is, X_t is equal to the mean square limit of the partial sums $\sum_{-n}^{n} \beta_j Z_{t-j}$. The proof is omitted here. Condition (1.5) also guarantees the convergence of the more general series (1.7) in the following theorem.

Proposition 1.1.5. *Suppose $\{Y_t\}$ is a stationary time series with $E[Y_t] = 0$ and autocovariance function γ_Y. If $\sum_{j=-\infty}^{\infty} |\beta_j| < \infty$, then the series*

$$X_t = \sum_{j=-\infty}^{\infty} \beta_j Y_{t-j} = \beta(B)Y_t \tag{1.7}$$

is stationary with mean 0 and autocovariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \beta_j \beta_k \gamma_Y(h+k-j) \tag{1.8}$$

If X_t is a linear process, then the ACVF is given by

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \beta_j \beta_k \sigma^2 \tag{1.9}$$

Proof. Note first that the same argument used in the case of the linear process above can be used to show the convergence of (1.7) by simply replacing σ with $\sqrt{\gamma_Y(0)}$. By assumption $E[Y_t] = 0$, so we may write

$$E[X_t] = E\left[\sum_{j=-\infty}^{\infty} \beta_j Y_{t-j}\right] = \sum_{j=-\infty}^{\infty} \beta_j E[Y_{t-j}] = 0.$$

Therefore we can write the ACVF as

$$\begin{aligned} E[X_{t+h}X_t] &= E\left[\left(\sum_{j=-\infty}^{\infty} \beta_j Y_{t+h-j}\right)\left(\sum_{k=-\infty}^{\infty} \beta_k Y_{t-k}\right)\right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \beta_j \beta_k E[Y_{t+h-j} Y_{t-k}] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \beta_j \beta_k \gamma_Y(h-j+k) \end{aligned}$$

which does not depend on t. In the foregoing equations, we have moved the expectation inside of the summation since the sequence $\{\beta_j\}$ is absolutely summable by assumption. Thus, $\{X_t\}$ meets the stationarity conditions with the autocovariance function given by (1.8). If $\{Y_t\} \sim \text{WN}(0, \sigma^2)$, then by definition $\{X_t\}$ is a linear process. In this special case we note that $\gamma_Y(h-j+k) = \sigma^2$ if $k = k - h$ and 0 otherwise, and thus (1.9) holds.

□

As a remark, one can show that it follows from the absolute convergence of (1.7) that linear filters with absolutely summable coefficients can be applied in succession to a stationary series $\{Y_t\}$

to generate a new stationary series $\{W_t\}$. Specifically, we can apply filters of the form

$$\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$$

and

$$\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$$

where in each case the coefficients are absolutely summable. Letting

$$\beta(B) = \alpha(B)\nu(B) \tag{1.10}$$

we can then apply the filters in succession to obtain

$$W_t = \beta(B)Y_t = \sum_{j=-\infty}^{\infty} \beta_j Y_{t-j}.$$

Here $\beta_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}$, and thus the order of applying the filters is interchangeable.

Armed with Proposition 1.1.5 we now return to the example of the AR(1) process given above. We wish show the existence of a unique solution to equation (1.2) under the given conditions. Again suppose that $\{X_t\}$ is a stationary solution of

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \ldots, \tag{1.11}$$

under the conditions $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$. To show the existence of such a solution consider the process

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}. \tag{1.12}$$

We can easily check by substitution that (1.12) is a solution of (1.11). Note that the sequence of the coefficients $\{\phi^j\}$ for $j \geq 0$ is absolutely summable since $|\phi| < 1$. By 1.1.5, the series $\{X_t\}$ is

stationary with mean 0, and its autocovariance function is given by

$$\gamma_X(h) = \sum_{j=0}^{\infty} \phi^h \phi^{j+h} \sigma^2 = \frac{\sigma^2 \phi^h}{1-\phi^2},$$

for $h \geq 0$, where we have used the formula for the sum of an infinite geometric series. It remains to be shown that (1.12) is the *unique* solution of (1.11). Suppose $\{Y_t\}$ is any stationary solution of (1.11). By iterating k times we find that

$$\begin{aligned} Y_t &= \phi Y_{t-1} + Z_t \\ &= Z_t + \phi Z_{t-1} + \phi^2 Y_{t-2} \\ &= \dots \\ &= Z_t + \phi Z_{t-1} + \cdots + \phi^k Z_{t-k} + \phi^{k+1} Y_{t-k-1}. \end{aligned}$$

Using this representation, we can show that Y_t is equal in mean square limit to $\sum_{j=0}^{\infty} \phi^j Z_{t-j}$. We note that since $\{Y_t\}$ is stationary by assumption, then $E[Y_t^2]$ must be finite and cannot depend on t, so we find that

$$\begin{aligned} E[(Y_t - \sum_{j=0}^{k} \phi^j Z_{t-j})^2] &= \phi^{2k+2} E[Y_{t-k-1}^2] \\ &\to 0 \end{aligned}$$

as $k \to \infty$ since $|\phi| < 1$. Thus we find that (1.12) is the unique stationary solution to (1.11) under the given conditions.

Let us explore the case where we assume that $|\phi| > 1$. It is clear that series given in (1.12) will not converge in this case. We begin by writing (1.11) in the slightly different form

$$\begin{aligned} X_{t+1} &= \phi X_t + Z_{t+1} \\ \implies X_t &= -\phi^{-1} Z_{t+1} + \phi^{-1} X_{t+1} \end{aligned}$$

Once again, iterating this expression k times gives

$$\begin{aligned} X_t &= -\phi^{-1}Z_{t+1} - \phi^{-2}Z_{t+2} + \phi^{-2}X_{t+2} \\ &= \dots \\ &= -\phi^{-1}Z_{t+1} - \cdots - \phi^{-k-1}Z_{t+k+1} + \phi^{-k-1}X_{t+k+1}. \end{aligned}$$

Reasoning as we did for the $|\phi| < 1$ case, we find that

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j} \tag{1.13}$$

is the unique stationary solution for (1.11) under the modified condition. It can be shown if $\phi = \pm 1$ that there is no stationary solution of (1.11).

We immediately notice that (1.13) is a fundamentally different type of solution than (1.12), in the sense that in (1.13) X_t is represented in terms of *future* values of $\{Z_t\}$, i.e. X_t depends on values of Z_s where $s > t$. As a result, the solution represented in (1.13) is often considered physically unrealistic. In contrast, for the solution in (1.12), X_t is uncorrelated with Z_s for all $s > t$. In this case, $\{X_t\}$ is represented by (1.11) in terms of $\{Z_s, s \leq t\}$ and depends only on the past and present values of Z_s. Here $\{X_t\}$ is referred to as a causal or future-independent function of $\{Z_t\}$, or we simply say that $\{X_t\}$ is a causal autoregressive process. From these considerations, it is conventional when seeking a stationary AR(1) model to consider only models where $|\phi| < 1$. We shall return to the concept of causality in a more general sense below.

To illustrate the utility of employing linear filters based on the backward shift operator, we note that in the case where $|\phi| < 1$ we can obtain the unique stationary solution (1.12) easily using equation (1.10). Let $\phi(B) = 1 - \phi B$ and $\kappa(B) = \sum_{j=o}^{\infty} \phi^j B^j$ and define

$$\beta(B) := \phi(B)\kappa(B) = 1.$$

If we apply $\kappa(B)$ to both sides of (1.11) we find that

$$X_t = \kappa(B)Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

which is the same expression as (1.12).

We now examine in greater detail one of the key classes of linear processes known as autoregressive moving-average or ARMA processes. We have already mentioned some of the most simple types of ARMA processes in the examples above. We will begin with the example of the ARMA(1,1) process. This will allow us to develop insights into some important properties of these models. We will then move onto higher-order ARMA(p, q) processes.

Example 1.1.6. ARMA(1,1) Process
A time series $\{X_t\}$ is an **ARMA(1,1) process** if it is stationary and satisfies the equation

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1} \tag{1.14}$$

for all t, where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and $\phi + \theta \neq 0$. We may write (1.14) more compactly by using the backward shift operator

$$\phi(B)X_t = \theta(B)Z_t \tag{1.15}$$

Here $\phi(B)$ and $\theta(B)$ represent the linear filters

$$\phi(B) = 1 - \phi B \text{ and } \phi(B) = 1 + \theta B.$$

We wish to find stationary solutions for (1.14). In light of the previous examples, we expect that the existence and form of stationary solutions will depend on the values of ϕ and θ. First suppose that $|\phi| < 1$. Let $\kappa(z)$ represent the power series expansion of $1/\phi(z)$

$$\kappa(z) = \sum_{j=0}^{\infty} \phi^j z^j \tag{1.16}$$

and note that the coeffecients are absolutely summable under the current assumption. Using equation (1.10) we have $\kappa(B)\phi(B) = 1$. Thus we can apply $\kappa(B)$ to both sides of (1.15) to obtain

$$X_t = \kappa(B)\theta(B)Z_t = \beta(B)Z_t. \tag{1.17}$$

To find a more explicit form for the stationary solution, we note that

$$\beta(B) = \sum_{j=0}^{\infty} \beta_j B^j = (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B).$$

To determine the coefficients β_j, we can multiply the two expressions on the right-hand side to find that $\beta_0 = 1$ and $\beta_j = (\beta + \theta)\beta^{j-1}$ for $j \geq 1$. Thus, applying the operator $\beta(B)$, equation (1.17) becomes

$$X_t = Z_t + (\beta + \theta) \sum_{j=1}^{\infty} Z_{t-j} \tag{1.18}$$

We can then use an argument similar to the one used in the example of the AR(1) process above to show that this is the unique stationary solution to (1.14). Such a process is known as an MA(∞) process.

Now consider the case where $|\phi| > 1$. It can be shown that $1/\phi(z)$ may be represented in terms of the powers of z^-1:

$$\frac{1}{\phi(z)} = -\sum_{j=1}^{\infty} \phi^{-j} z^{-j}.$$

Again, the coefficients are absolutely summable since $|\phi| > 1$ by assumption. We may now argue along the same lines as in the previous case to find the unique stationary solution to (1.14). Let $\kappa(B) = -\sum_{j=1}^{\infty} \phi^{-j} B^{-j}$ and again note that $\kappa(B)\phi(B) = 1$. Applying $\kappa(B)$ to both sides of (1.15) and carefully multiplying the terms yields

$$\begin{aligned} X_t &= \kappa(B)\theta(B)Z_t \\ &= -\theta\phi^{-1}Z_t - (\theta + \phi) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t+j} \end{aligned} \tag{1.19}$$

which is the unique stationary solution. As in the example of the AR(1), there is no stationary solution to (1.14) if $\phi = \pm 1$. Note that by our definition an ARMA(1,1) process is stationary and thus there can be no ARMA(1,1) processes where $\phi = \pm 1$.

When $|\phi| < 1$, the solution given by (1.18) shows that $\{X_t\}$ is a causal function of $\{Z_t\}$, since X_t can be expressed in terms of the past and present values Z_s, where $s \leq t$. When $|\phi| > 1$, the solution given by (1.19) reveals that X_t depends on the current and future values of Z_s where $s \geq t$ and thus $\{X_t\}$ is referred to as a noncausal function of $\{Z_t\}$.

Closely related to the notion of causality is the concept of invertibility. We refer to $\{X_t\}$

as invertible if Z_t can be expressed in terms of X_s where $s \leq t$. We can readily show that the ARMA(1,1) process above is invertible if $|\theta| < 1$ by employing an argument similar to that used to find the stationary solutions above. Let $\tau(z)$ be the power series expansion of $1/\theta(z)$:

$$\tau(z) = \sum_{j=0}^{\infty} (-\theta)^j z^j.$$

and note that under our assumption the series has absolutely summable coefficients. Using equation (1.10), we have $\tau(B)\theta(B) = 1$. We may now apply $\tau(B)$ to both sides of (1.15) to obtain

$$Z_t = \tau(B)\phi(B)X_t = \omega(B)X_t.$$

Here $\omega(B)$ takes the form

$$\omega(B) = \sum_{j=0}^{\infty} \omega_j B^j = (1 - \theta B + (-\theta)^2 B^2 + \dots)(1 - \phi B).$$

After multiplying the right hand side and gathering therms we arrive at

$$Z_t = X_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{t-j}.$$

which shows that the ARMA(1,1) process is invertible if $|\theta| < 1$, since Z_t depends on current and past values of X_s where $s \leq t$. In much the same way that we showed that $\{X_t\}$ is noncausal when $|\phi| > 1$, we can also show that $\{X_t\}$ is noninvertible if $|\theta| > 1$. Through an analogous argument we find

$$Z_t = -\phi\theta^{-1} X_t + (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{-j-1} X_{t+j}.$$

Thus we see that if $|\theta| > 1$ then the process is noninvertible since since Z_t is expressed in terms of current and future values of X_s where $s \geq t$.

We now present the more general class of ARMA(p, q) models. Having discussed the ARMA(1,1) process and examined the conditions under which it has the important properties of causality and invertibility, we may now develop these concepts for the broader class of models.

Definition 1.1.1. A process $\{X_t\}$ is an **ARMA(p, q) process** if it is weakly stationary and satisfies

the equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{1.20}$$

for every t where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. A process $\{X_t\}$ is an ARMA(p, q) with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process. We can write (1.20) more concisely as

$$\phi(B)X_t = \theta(B)Z_t \tag{1.21}$$

where

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

and B is the backward shift operator. If $\phi(z) = 1$, then the series $\{X_t\}$ is known as an **MA(q)** or **moving-average process of order q**. If $\theta(z) = 1$, then $\{X_t\}$ is known as an **AR(p)** or **autoregressive process of order p**.

For the example of the ARMA(1,1) process, we noted that a stationary solution exists if and only if $|\phi| \neq 1$. This condition is the same as requiring that the polynomial $\phi(z) = 1 - \phi_1 z \neq 0$ for $z = \pm 1$. We can prove an analogous condition for the ARMA(p, q) process using the polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$. For completeness, we will provide the proofs of the conditions for causality, invertibility, and stationarity for the general ARMA(p, q) process. The following proofs follow the general method given by Brockwell and Davis in *Time Series: Theory and Methods* [3]. We begin by giving the formal defintions of causality and invertibility.

Definition 1.1.2. An ARMA(p, q) process $\{X_t\}$ is **causal**, or a **causal function** of $\{Z_t\}$, if there exists a sequence of constants $\{\beta_t\}$ such that $\sum_{j=0}^{\infty} |\beta_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j} \tag{1.22}$$

for all t.

Definition 1.1.3. An ARMA(p, q) process $\{X_t\}$ is **invertible** if there exists a sequence of constants

$\{\kappa_t\}$ such that $\sum_{j=0}^{\infty} |\kappa_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \kappa_j X_{t-j} \tag{1.23}$$

for all t.

The following theorem provides necessary and sufficient conditions for causality of an ARMA process and gives an explicit representation of the coefficients in (1.22).

Theorem 1.1.7. *Let $\{X_t\}$ be an ARMA(p, q) process for which the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeroes. Then $\{X_t\}$ is causal if and only if $\phi(z) \neq 0$ for all complex z such that $|z| \leq 1$. The coefficients in* (1.22) *are determined by*

$$\beta(z) = \sum_{j=0}^{\infty} \beta_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1 \tag{1.24}$$

and may be computed by recursion relations.

Proof. Assume first that $\phi(z) \neq 0$ if $|z| \leq 1$. There must exist an $\epsilon > 0$ such that $1/\phi(z)$ can be represented by the power series

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \lambda_j z^j = \lambda(z), \quad |z| < 1 + \epsilon. \tag{1.25}$$

Thus $\lambda_j(1 + \epsilon/2)^j \to 0$ as $j \to \infty$. Therefore, there must exist $K \in (0, \infty)$ such that

$$|\lambda_j| < K(1 + \epsilon/2)^{-j} \quad \text{for all } j = 0, 1, 2, \ldots$$

Hence we have that the coefficients in (1.25) are absolutely summable, i.e, $\sum_{j=0}^{\infty} |\lambda_j| < \infty$. Noting that $\lambda(z)\phi(z) = 1$ for $|z| \leq 1$, by Proposition (1.1.5) we can apply the operator $\lambda(B)$ to both sides of (1.21) which yields

$$\lambda(B)\phi(B)X_t = X_t = \lambda(B)\theta(B)Z_t.$$

This gives the causal representation

$$X_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j}$$

with the coefficients $\{\beta_j\}$ determined by (1.24).

Now assume that $\{X_t\}$ is causal and thus has the representation given in equation (1.22).

Hence we may represent X_t as

$$X_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j} = \beta(B) Z_t$$

where the sequence of coefficients $\{\beta_j\}$ is absolutely summable, and so using (1.21) we may write

$$\theta(B) Z_t = \phi(B) X_t = \phi(B)\beta(B) Z_t. \tag{1.26}$$

Now define $\upsilon(z) := \phi(z)\beta(z) = \sum_{j=0}^{\infty} \upsilon_j z^j$ where $|z| \leq 1$. We can write (1.26) as

$$\sum_{j=0}^{\infty} \theta_j Z_{t-j} = \sum_{j=0}^{\infty} \upsilon_j Z_{t-j}.$$

Noting that $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and thus is an uncorrelated sequence, we can determine the coefficients υ_k by taking the inner product of each side of the equation with Z_{t-k}. This yields that $\upsilon_k = \theta_k$ for $k = 0, 1, \ldots, q$ and $\upsilon_k = 0$ for $k > q$. Thus the two polynomials $\theta(z)$ and $\upsilon(z)$ are equal and we have

$$\theta(z) = \upsilon(z) = \phi(z)\beta(z) \quad |z| < 1.$$

By assumption $\theta(z)$ and $\phi(z)$ have no common zeroes, and noting that $|\beta(z)| < \infty$ for $|z| \leq 1$, we see that $\phi(z)$ cannot not be zero for $|z| \leq 1$, thus proving the result.

□

Having specified the necessary and sufficient conditions for causality of an ARMA(p, q) process, we can prove a similar result regarding the invertibility conditions. We can also determine the coefficients in the representation in (1.23).

Theorem 1.1.8. *Let $\{X_t\}$ be an ARMA(p, q) process for which the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeroes. $\{X_t\}$ is invertible if and only if $\theta(z) \neq 0$ for all complex z such that $|z| \leq 1$. The coefficients in* (1.23) *are determined by*

$$\kappa(z) = \sum_{j=0}^{\infty} \kappa_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1 \tag{1.27}$$

and may be computed by recursion relations.

Proof. We proceed in a manner similar to the previous proof. We begin by assuming that $\theta(z) \neq 0$

if $|z| \leq 1$. There must exist an $\epsilon > 0$ such that $1/\theta(z)$ can be represented by the power series

$$\frac{1}{\theta(z)} = \sum_{j=0}^{\infty} \lambda_j z^j = \lambda(z), \quad |z| < 1 + \epsilon. \tag{1.28}$$

Thus $\lambda_j(1 + \epsilon/2)^j \to 0$ as $j \to \infty$. Therefore, there must exist $K \in (0, \infty)$ such that

$$|\lambda_j| < K(1 + \epsilon/2)^{-j} \quad \text{for all } j = 0, 1, 2, \ldots$$

and we may conclude that the coefficients in (1.28) are absolutely summable, i.e, $\sum_{j=0}^{\infty} |\lambda_j| < \infty$. Noting that $\lambda(z)\theta(z) = 1$ for $|z| \leq 1$, by Proposition (1.1.5) we can apply the operator $\lambda(B)$ to both sides of (1.21) which yields

$$\lambda(B)\phi(B)X_t = \lambda(B)\theta(B)Z_t = Z_t.$$

Thus we obtain the representation given in (1.23)

$$Z_t = \sum_{j=0}^{\infty} \kappa_j X_{t-j}$$

with the coefficients $\{\beta_j\}$ determined by (1.27).

To prove the converse, assume that $\{X_t\}$ is invertible and therefore it can be represented as $Z_t = \sum_{j=0}^{\infty} \kappa_j X_{t-j}$ where the sequence of coefficients $\{\kappa_j\}$ is absolutely summable. Using (1.21) we have

$$\phi(B)Z_t = \kappa(B)\phi(B)X_t = \kappa(B)\theta(B)Z_t. \tag{1.29}$$

Now define $\upsilon(z) := \kappa(z)\theta(z) = \sum_{j=0}^{\infty} \upsilon_j z^j$ where $|z| \leq 1$. We can write (1.29) as

$$\sum_{j=0}^{\infty} \phi_j Z_{t-j} = \sum_{j=0}^{\infty} \upsilon_j Z_{t-j}.$$

Noting again that $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and thus is an uncorrelated sequence, we can determine the coefficients υ_k by taking the inner product of each side of the equation with Z_{t-k}. This yields that $\upsilon_k = \phi_k$ for $k = 0, 1, \ldots, p$ and $\upsilon_k = 0$ for $k > p$. Thus the two polynomials $\phi(z)$ and $\upsilon(z)$ are equal and we have

$$\phi(z) = \upsilon(z) = \kappa(z)\theta(z) \quad |z| < 1.$$

By assumption $\theta(z)$ and $\phi(z)$ have no common zeroes, and noting that $|\kappa(z)| < \infty$ for $|z| \leq 1$, we see that $\theta(z)$ cannot not be zero for $|z| \leq 1$, which was the desired result. $\square$

-

We will finish our overview of ARMA(p, q) processes by considering the conditions under which a unique stationary solution exists for the general ARMA equations given by (1.20). Here we will not assume that $\{X_t\}$ is causal or invertible. From considerations similar to those in the preceding proofs, it can be shown that if $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeroes, and if $\phi(z) = 0$ for some complex $|z| = 1$, then there can be no stationary solution of (1.20). However, if $\phi(z) \neq 0$ for all complex $|z| = 1$, then it is known from complex analysis that there must exist an $r > 1$ such that

$$\theta(z)\phi(z)^{-1} = \sum_{-\infty}^{\infty} \beta_j z^j = \beta(z) \qquad r^{-1} < |z| < r \tag{1.30}$$

where the series is absolutely convergent over the designated region. We will make use of this well-known result in the following proof.

Theorem 1.1.9. *If $\phi(z) \neq 0$ for all complex z such that $|z| = 1$, then the equation for the ARMA process given by $\phi(B)X_t = \theta(B)Z_t$ will have a unique solution of the form*

$$X_t = \sum_{-\infty}^{\infty} \beta_j Z_{t-j} \tag{1.31}$$

with the coefficients determined by equation (1.30).

Proof. Since $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, by Proposition (1.1.5) a solution of the form in (1.31) will be a stationary process. Now we apply the operator $\phi(B)$ to both sides of (1.31) which yields

$$\phi(B)X_t = \phi(B)\beta(B)Z_t = \theta(B)Z_t$$

where we have again employed Proposition (1.1.5). Thus we find that $\{X_t\}$ is a stationary solution of the ARMA process equations.

Now suppose that $\{X_t\}$ is a stationary solution of $\phi(B)X_t = \theta(B)Z_t$. By assumption $\phi(z) \neq 0$ for all complex z such that $|z| = 1$, and by the complex analysis result above there must exist $\delta > 1$ such that $\phi(z)^{-1} = \sum_{-\infty}^{\infty} \lambda_j z^j = \lambda(z)$ is absolutely convergent over the region

$-\delta < |z| < \delta$. Thus we may apply $\lambda(B)$ of $\phi(B)X_t = \theta(B)Z_t$ to obtain

$$\lambda(B)\phi(B)X_t = \lambda(B)\theta(B)Z_t$$

which we rewrite as

$$X_t = \beta(B)Z_t$$

where $\beta(B) = \lambda(B)\theta(B)$. □

This concludes the overview of ARMA(p, q) processes and the conditions pertaining to the key properties of stationarity, causality, and invertibility. In the next section, we will examine maximum likelihood estimation and some important results that provide context for our work in Chapter 2.

1.2 Maximum Likelihood Estimation

One of the most popular and powerful techniques for parameter estimation is the method of maximum likelihood estimation. If we begin with a random sample, by maximizing a likelihood function, we can arrive at estimators for the parameters of the assumed underlying model such that the observed data sample is most probable. These estimators are referred to as maximum likelihood estimators or MLE's. In this section we will examine the likelihood function, and several simple examples of maximum likelihood estimators. We will also consider some key properties of MLEs. The actual method of maximizing the likelihood function depends on the form of the function itself. If the likelihood function is easy to differentiate, then we can apply the derivative test to determine its maximum. In practical situations however, it will often be necessary to employ numerical methods to find the maximum. In the discussion that follows, we adopt a frequentist approach and follow the general development of these topics given by Casella and Berger in *Statistical Inference* and Rossi in *Mathematical Statistics : An Introduction to Likelihood Based Inference* [6, 28].

We begin by defining the likelihood function. We will assume that all random samples are iid draws from an underlying distribution.

Definition 1.2.1. Let $X_1, \ldots, X_n$ be a random sample from a population with pdf or pmf $f(x|\theta_1, \ldots, \theta_k)$.

The **likelihood function** is defined as

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = L(\theta_1, \ldots, \theta_k|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k).$$

In maximum likelihood estimation, the objective is to find the values of the parameters that maximize the function $L(\boldsymbol{\theta}|\boldsymbol{x})$ over the parameter space, i.e. we want to find

$$\hat{\theta} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}|\boldsymbol{x}).$$

Here Θ is a finite-dimensional subspace of $\mathbb{R}^n$ called the parameter space.

Definition 1.2.2. Let $\mathbf{x}$ be a point in the sample space, and let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $L(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum value as a function of θ for fixed $\mathbf{x}$. For a sample $\mathbf{X}$, $\hat{\theta}(\mathbf{X})$ is a **maximum likelihood estimator** of the parameter θ.

For a specific set of observations, the particular value $\hat{\theta}$ that maximizes $L(\boldsymbol{\theta}|\boldsymbol{x})$ as a function of θ is the maximum likelihood estimate. In order for the estimator to exist, a sufficient but not necessary condition is that the likelihood function is continuous over a compact parameter space. If the parameter space is open, it is possible that the likelihood never reaches a supremum value on that space.

Maximum likelihood estimation is subject to two potential difficulties that arise in finding the maximum of a function. First, we must be able to find the maximum, and then we must verify that the maximum is global rather than local. This can often be accomplished through well-known techniques from differential calculus, although even for some common densities the problem is not straightforward. Numerical sensitivity is the second major issue that can arise. Its important to be aware of the degree to which the estimate is sensitive to small changes in the data. Caution should be exercised when using an MLE if we find that small changes in the sample produce drastically different estimates.

If the likelihood function is differentiable with respect to θ_i, then we can identify possible candidates for the MLE by finding the values of $(\theta_1, \ldots, \theta_k)$ that are solutions to the equations

$$\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{x})}{\partial \theta_i} \qquad i = 1, \ldots, k. \tag{1.32}$$

The solutions to (1.32) are not necessarily MLEs since a first derivative of 0 is a necessary but not sufficient condition for a maximum of a function to occur on the interior of its domain. Recall that if the first derivative at a point is zero, the point may be a local maximum or minimum, a global maximum or minimum, or a critical point. We are interested in finding the global maximum, and thus we must also check the endpoints of the domain where the first derivative is not necessarily 0.

Sometimes when performing maximum likelihood estimation, it is necessary to impose additional constraints on the allowed values of the parameters. In such cases, the restricted parameter space can be expressed as

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^k, h(\boldsymbol{\theta}) = 0\}.$$

Here $h(\boldsymbol{\theta}) = [h_1(\boldsymbol{\theta}), \ldots, h_r(\boldsymbol{\theta})]$ is a function from $\mathbb{R}^k$ to $\mathbb{R}^r$. Thus, to find the MLE of $\boldsymbol{\theta}$, we need to maximize the likelihood function subject to the constraint $h(\boldsymbol{\theta}) = 0$. In most practical situations, the method of Lagrange may be used to impose the constraints.

When finding MLEs, it is often convenient to work with the natural logarithm of the likelihood function. This function, known as the **log likelihood** is frequently much easier to differentiate than the likelihood function, and is denoted by

$$\ell(\boldsymbol{\theta}|\boldsymbol{x}) := \ln L(\boldsymbol{\theta}|\boldsymbol{x}).$$

The natural logarithm function is monotonically increasing, which conveniently guarantees that the maximum of $\ell(\boldsymbol{\theta}|\boldsymbol{x})$ occurs at the same value of $\boldsymbol{\theta}$ as the maximum of $L(\boldsymbol{\theta}|\boldsymbol{x})$. We now proceed with some examples to illustrate the process of finding MLEs for some common distributions.

Example 1.2.1. Bernoulli MLE

Let $X_1, \ldots, X_n$ be a randnom sample from a Bernoulli(p) distribution. The likelihood function is given by

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^y(1-p)^{1-y}$$

where $y = \sum_{i=1}^{n} x_i$. After taking the natural log, we find the log likelihood is easier to differentiate:

$$\ell(p|\boldsymbol{x}) = y \ln p + (n-y)\ln(1-p). \tag{1.33}$$

Differentiating and setting (1.33) equal to 0 , we find that for $0 < y < n$

$$\begin{aligned}\frac{d\ell(p|\boldsymbol{x})}{dp} &= \frac{dy\ln p}{dp} + \frac{d(n-y)\ln(1-p)}{dp} \\ &= \frac{y}{p} + (n-y)\left(\frac{-1}{1-p}\right) \\ &= 0\end{aligned}$$

which gives the solution $\hat{(p)} = y/n = \sum_{i=1}^{n} x_i/n$. To check that this is a maximum we can find the second derivative

$$\frac{d^2\ell(\boldsymbol{\theta}|\boldsymbol{x})}{d^2p} = \frac{-y}{p^2} - (n-y)\left(\frac{1}{(1-p)^2}\right)$$

which is negative since $p \in [0,1]$ and $x_i \in 0, 1$. Thus by the second derivative y/n is a maximum. By checking the endpoints, we can easily verify that this a global maximum. If $y = 0$ or $y = n$, the log likelihood is given by

$$\ell(p|\boldsymbol{x}) = \begin{cases} n\ln(1-p) & \text{if} y = 0 \\ n\ln p & \text{if} y = n. \end{cases}$$

Note that in either case $\ell(\boldsymbol{\theta}|\boldsymbol{x})$ is a monotone function of p, and by setting the first derivative equal to zero we again find that $\hat{p} = y/n$ for both cases. Hence, $\sum_{i=1}^{n} X_i/n$ is the MLE of p. Before moving on to examples involving continuous distributions, we examine another MLE for a discrete distribution.

Example 1.2.2. Binomial MLE with unknown k

Let $X_1, \ldots, X_n$ be a random sample from a Binomial(k, p) distribution. The likelihood function is given by

$$L(k|\boldsymbol{x}, p) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i}(1-p)^{k-x_i}.$$

Suppose here that p is known and k is unknown. This would be the situation if we were flipping a fair coin ($p = 1/2$) and we observe x_i heads but we do not know how many coin flips occurred. Maximizing $L(k|\boldsymbol{x}, p)$ with respect to k is not straightforward due to the presence of factorials, plus we require that k must be an integer. Here we employ a different approach to find the MLE for k.

First we note that $L(k|\boldsymbol{x}, p) = 0$ if $k < \max_i x_i$. Hence the MLE must be an integer k such

that $k \geq max_i x_i$. In order to maximize $L(k|\boldsymbol{x}, p) = 0$, this integer k must satisfy

$$\frac{L(k-1|\boldsymbol{x}, p)}{L(k|\boldsymbol{x}, p)} \geq 1 \quad \text{and} \quad \frac{L(k|\boldsymbol{x}, p)}{L(k+1|\boldsymbol{x}, p)} < 1. \tag{1.34}$$

In order to establish that there is only one such k, consider the ratio of the likelihoods given by

$$\frac{L(k|\boldsymbol{x}, p)}{L(k-1|\boldsymbol{x}, p)} = \frac{(k(1-p))^n}{\prod_{i=1}^n (k - x_i)}.$$

Combining this expression with (1.34), the condition for a maximum becomes

$$(k(1-p))^n \geq \prod_{i=1}^n (k - x_i) \quad \text{and} \quad ((k+1)(1-p))^n < \prod_{i=1}^n (k + 1 - x_i).$$

If we divide by k^n and let $z = 1/k$, we seek a solution of

$$(1-p)^n = \prod_{i=1}^n (1 - x_i z) \tag{1.35}$$

where $0 \leq z \leq 1/\max_i x_i$. We note that the right-hand side is a strictly decreasing function of z in the given range. The right-hand side will equal unity when $z = 0$ and zero when $z = 1/\max_i x_i$. Hence there must be a unique solution $\hat{z}$ to the equation (1.35). Although $1/\hat{z}$ may not be an integer, the k which is the largest integer less than or equal to $1/\hat{z}$ will satisfy the inequalities above, and is thus the MLE. We can find this unique maximum for the likelihood function numerically by solving an nth degree polynomial quantity. From this example we see that even for some common distributions, finding the maximum of the likelihood function with differential calculus is not straightforward, and we must consider other methods.

Example 1.2.3. Exponential MLE

As an example of maximum likelihood estimation for a common continuous distribution, consider a random sample $X_1, \ldots, X_n$ from an exponential distribution with unknown parameter $\lambda > 0$. The likelihood function is given by

$$L(\lambda|\boldsymbol{x}) = \prod_{i=1}^n \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right).$$

Taking a natural log yields

$$\begin{aligned}\ell(\lambda|\boldsymbol{x}) &= \ln\left(\lambda^n \exp\left(-\lambda\sum_{i=1}^{n} x_i\right)\right) \\ &= \ln(\lambda^n) + \ln\left(\exp\left(-\lambda\sum_{i=1}^{n} x_i\right)\right) \\ &= n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i.\end{aligned}$$

Differentiating this expression with respect to λ and setting the result to zero, we find that

$$\frac{d\ell(p|\boldsymbol{x})}{dp} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0.$$

Rearranging, we find that

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

It is easy to show that this solution is the global maximum of the likelihood function for $\lambda > 0$. Thus $\hat{\lambda}$ is the MLE. Note that this is simply the reciprocal of the sample mean, which is intuitive since $E[X] = 1/\lambda$ for $X \sim \text{Exp}\ (\lambda)$.

Next, we consider an example of maximum likelihood estimation with two unknown parameters.

Example 1.2.4. Normal MLE with unknown μ and σ^2

Suppose $X_1, \ldots, X_n$ is a random sample such that $X_i \sim \mathcal{N}(\mu, \sigma^2)$ for $i = 1, \ldots, n$, where both parameters μ and $\sigma^2 > 0$ are unknown. The likelihood is

$$L(\mu, \sigma^2|\boldsymbol{x}) = \frac{1}{(2\kappa\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

This expression is differentiable; moreover, taking natural logs gives

$$\ell(\mu, \sigma^2|\boldsymbol{x}) = -\frac{n}{2}\ln 2\kappa - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^2}.$$

Since there are two unknown parameters in this case, we take the partial derivative of $\ell(\mu, \sigma^2|\boldsymbol{x})$

with respect to each parameter, which gives

$$\frac{\partial \ell(\mu, \sigma^2|\boldsymbol{x})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

and

$$\frac{\partial \ell(\mu, \sigma^2|\boldsymbol{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Setting each partial derivatives to zero and solving for the respective parameters, we find the estimators

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

and

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}.$$

It remains to verify that the solution $(\hat{\mu}, \hat{\sigma^2})$ is a global maximum. We can simplify the process by noting that if $\mu \neq \bar{x}$, then $\sum_{i=1}^{n}(x_i - \mu)^2 > \sum_{i=1}^{n}(x_i - \bar{x})^2$, which implies that

$$\frac{1}{(2\kappa\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right) \geq \frac{1}{(2\kappa\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

Thus, the estimator $\bar{x}$ already maximizes the likelihood with respect to μ. The problem of showing that $(\hat{\mu}, \hat{\sigma^2})$ is a global maximum is now reduced to the one-dimensional problem of showing that the left-hand side of the above inequality achieves a global maximum at

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}.$$

We can then handle this readily with univariate calculus techniques. In particular, we verify that

$$(\hat{\mu}, \hat{\sigma^2}) = \left(\bar{X}, \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right)$$

are the MLEs for (μ, σ^2).

When dealing with finite sample sizes, other estimators may frequently be closer than the MLE to the true parameter value. However, maximum likelihood estimators possess several useful limiting properties. Under fairly broad conditions, the MLE is a consistent estimator, i.e. as the

sample size goes to infinity the sequence of MLEs converges in probability to the true parameter:

$$\hat{\theta}_{MLE} \xrightarrow{p} \theta.$$

In addition, in the limit as sample size goes to infinity, the MLE is functionally equivariant. This means that for a parameter θ, if $g(\theta)$ is a function of θ, then $\hat{\tau} = g(\hat{\theta}_{MLE})$ is the MLE of $g(\theta)$.

Furthermore, the MLE is an asymptotically efficient estimator of the parameter in question, that is it achieves the Cramér–Rao lower bound as the sample size increases to infinity. In particular, under certain regularity conditions,

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1})$$

where $\mathcal{I}$ is the Fisher information matrix, and thus the MLE is asymptotically normal.

Having considered some basic examples and properties of maximum likelihood estimators, we will now give a brief overview of Lindley processes.

1.3 Introduction to Lindley Processes

A Lindley random walk, also known as a Lindley recursion or storage model, is a staple model that governs customer waiting times in a simple single server queue over time [17, 20]. A Lindley process is also an appropriate mass balance model for some climate data, including snow depths, frozen soil depths, and ice thicknesses. A Lindley random walk $\{X_n\}_{n=0}^{\infty}$ obeys the recursion

$$X_n = \max(X_{n-1} + C_n, 0), \quad n \geq 1, \tag{1.36}$$

starting from some initial level X_0. The process $\{C_n\}_{n=1}^{\infty}$ is a sequence of random variables that we call the change process.

In classic G/G/1 queueing applications, the virtual waiting time of the nth customer (the time the customer waits until their service begins), denoted by W_n, obeys the Lindley recursion

$$W_n = \max(W_{n-1} + S_{n-1} - I_n, 0), \tag{1.37}$$

where I_n is the interarrival time between the $(n-1)$st and nth arriving customers and S_n is the service time of the nth customer. The typical assumptions in queueing theory are that $\{I_n\}_{n=1}^{\infty}$ and $\{S_n\}_{n=0}^{\infty}$ are independent and identically distributed (IID) non-negative random sequences. In this case, $C_n = S_{n-1} - I_n$ is IID. In a different vein of applications, [33] models daily snow depths at day n as the snow depth on the ground yesterday, plus any new snowfall, minus any meltoff or compaction since yesterday. To prevent non-identifiability, one amalgamates any new snow, meltoff, and compaction into a single change variable C_n and the model in (1.36) arises. Since daily weather is highly correlated, one would want a temporally correlated $\{C_n\}$. In queueing applications, the server may be inclined to work faster when there are more customers in the queueing, inducing correlation in $\{C_n\}$.

As storage levels cannot be negative, a Lindley random walk enforces a "hard boundary" at zero. One could transform Lindley random walk data prior to modeling, analyzing say $\ln(X_n + \epsilon)$ for some small $\epsilon > 0$ as a quantity taking on any real value. While this is not necessarily a bad idea, such a scheme cannot enforce the hard boundary of zero once the logarithm is undone. Indeed, the probability of an empty store, a key quantity for store administrators, would always be estimated as zero under this scheme. This issue makes it preferable to model the storage level directly (without transformation).

The vast amount of research on Lindley random walks has been done in queueing contexts that assume IID $\{C_n\}$. A classic reference here is [20]. In the correlated case, stability of waiting times for stationary $\{S_{n-1}, I_n\}$ were examined in [23, 22]. Discrete-time queues characterized by a discrete AR(1) processes were studied in [15] and [16]. Dependence can appear in a variety of queueing contexts, for example with batching or when processing multiple, distinct classes of customers, see [8, 9, 13, 14, 18, 29].

The purpose of this dissertation is to examine the Lindley random walk when $\{C_n\}$ is strictly stationary, and more specifically, a pth order autoregression. We provide a streamlined proof of stochastic stability that uses stochastic monotonicity and coupling techniques for technical efficiency. When $\{C_n\}$ is a pth order autoregression, the Markov properties (and lack thereof) of the Lindley walk are established. The model's statistical likelihood is then derived and novel particle filtering inference techniques are considered to estimate model parameters via likelihood methods.

The rest of this dissertation proceeds as follows. Section 2.1 establishes the ergodicity and stationarity of $\{X_n\}$ when $\{C_n\}$ is strictly stationary. Section 2.2 then establishes several Markov

structures for the process, even though $\{X_n\}$ is not a Markov chain itself. The final two sections of Chapter 2 move to statistical inference issues for the walk. In particular, Section 2.3 derives the model's likelihood function when $\{C_n\}$ is a pth order Gaussian autoregression. As the obtained likelihood function involves some unwieldy multivariate integrals, a particle filtering approach is devised to evaluate and optimize it in Section 2.5. A simulation study is given that demonstrates the accuracy of the approach. The dissertation concludes with discussion and comments in Chapter 3.

Chapter 2

Results

2.1 Stability and Stationarity

This section assumes that $\{C_n\}_{n=1}^{\infty}$ is strictly stationary. We want to show that the Lindley random walk is stable whenever $E[C_1] < 0$. While some old and recent literature also establishes these properties [23, 22], we present a simple elementary argument.

We first assume that $X_0 = 0$ and define $F_n(x) = P(X_n \leq x)$. To see that X_n is stochastically increasing in n, manipulations with (1.36) provide

$$\begin{aligned} X_n &= S_n - \min_{0\leq j\leq n} S_j \\ &= \max_{0\leq j\leq n} (S_n - S_j) \\ &= \max(0, C_n, C_n + C_{n-1}, \ldots, C_n + \cdots + C_1), \end{aligned}$$

where $S_n = C_1 + \cdots + C_n$.

Using this gives, for $x \geq 0$,

$$\begin{aligned} F_{n+1}(x) &= P(C_1 + C_2 + \cdots + C_{n+1} \leq x; C_2 + \cdots + C_{n+1} \leq x; \ldots; C_{n+1} \leq x) \\ &\leq P(C_2 + \cdots + C_{n+1} \leq x; C_3 + \cdots + C_{n+1} \leq x; \ldots; C_{n+1} \leq x) \\ &= P(C_1 + \cdots + C_n \leq x; C_2 + \cdots + C_n \leq x; \ldots; C_n \leq x) \\ &= F_n(x). \end{aligned}$$

Here, the inequality follows by dropping off the first terms in the probability and the second to last equality follows from strict stationary of $\{C_n\}$. This shows that $F_n(x)$ is monotone non-increasing in n for each fixed $x \geq 0$.

From this monotonicity, we may define $F_\infty(x) = \lim_{n\to\infty} F_n(x)$. It is clear that $F_\infty(x)$ is non-decreasing in x and takes values in [0,1]. To see that $F_\infty(x)$ is also right continuous in x, and hence a distribution function, note that

$$\lim_{h\downarrow 0} F_\infty(x+h) = \lim_{h\downarrow 0} \lim_{n\to\infty} F_n(x+h) = \lim_{n\to\infty} \lim_{h\downarrow 0} F_n(x+h) = \lim_{n\to\infty} F_n(x) = F_\infty(x).$$

The interchange of limit orders is justified by the fact that $F_n(x+h)$ is non-increasing with increasing n and decreasing h.

To show that $F_\infty(\cdot)$ is a proper cumulative distribution function (not vague), we need to make the law of large numbers (ergodic) assumption that

$$\frac{C_1 + \cdots + C_n}{n} \longrightarrow \mu_C \tag{2.1}$$

with probability one as $n \to \infty$ and that $\mu_C < 0$. A negative μ_C is necessary to induce a stable random walk even when $\{C_n\}$ is IID.

Now on any point in the probability space where (2.1) holds, $C_1 + \cdots + C_n \to -\infty$ as $n \to \infty$. On this path of $\{C_n\}_{n=1}^\infty$, we must have that

$$X_n = \max(0, C_n, C_n + C_{n-1}, \ldots, C_n + C_{n-1} + \cdots + C_1) \tag{2.2}$$

is bounded away from positive infinity in n. An implication of this is that $\limsup_{n\to\infty} X_n < \infty$, implying that X_∞, a random variable having the CDF $F_\infty(\cdot)$, is finite with probability one.

When $\mu_C = 0$, a proper limiting distribution may or may not arise. The case where $C_n \equiv 0$ provides an example where a trivial degenerate limiting distribution exits; when $\{C_n\}$ is IID and double exponentially distributed, $\{X_n\}_{n=0}^\infty$ is a null recurrent Markov chain. When $\mu_C > 0$, X_n will converge to infinity almost surely as $n \to \infty$ (assuming that $\{C_n\}$ obeys a law of large numbers) and no proper limiting distribution exists. See [23] for more on these cases, or any queueing text when $\{C_n\}$ is IID.

We now move to cases where the initial condition does not take X_0 as zero. For a fixed

sample path of $\{C_n\}_{n=1}^{\infty}$, let $\{X_n\}_{n=0}^{\infty}$ denote the process when $X_0 = 0$ and let $\{X_n^*\}_{n=0}^{\infty}$ denote the process when with initial condition is X_0^*. Here, X_0^* may or may not be random and both $\{X_n\}_{n=0}^{\infty}$ and $\{X_n^*\}_{n=0}^{\infty}$ are driven by the same sample path of $\{C_n\}_{n=1}^{\infty}$. Define the coupling time $T = \inf\{n \geq 0 : X_n = X_n^*\}$. Since the walk is pathwise ordered, $X_n \leq X_n^*$ for all $n \geq 0$. Thus, once $\{X_n^*\}$ first hits state zero, $X_n = X_n^*$ for all n thereafter; that is,

$$T \leq \inf\{n \geq 0 : X_n^* = 0\}.$$

Applying the classic coupling inequality [21] gives

$$\sup_A \mid P(X_n \in A) - P(X_n^* \in A) \mid \leq P(T_0 > n),$$

where the supremum is taken over all Borel measurable subsets of $[0, \infty)$ and $T_0 = \inf\{n \geq 0 : X_n^* = 0\}$.

An implication of the above is that if T_0 is a proper random variable, the limiting distribution will not depend on the initial content. We comment that no Markov property of the storage process has been used; our results all follow from stochastic monotonicity.

To see that T_0 is proper, note that

$$\begin{aligned} P(T_0 > n) &= P(X_t^* > 0 \text{ for all } t \text{ in } \{0, 1, \ldots, n\}) \\ &= P(X_0^* + C_1 + \cdots + C_t > 0 \text{ for all } t \text{ in } \{0, 1, \ldots, n\}) \\ &\leq P(X_0^* + C_1 + \cdots + C_n > 0). \end{aligned}$$

However, by (2.1), $C_1 + \cdots + C_n \to -\infty$ as $n \to \infty$ with probability one. Thus, T_0 is finite with probability one and the limit distribution does not depend on the initial condition X_0^*.

Next, we investigate the stationarity of $\{X_n\}_{n=0}^{\infty}$ when X_0 has its limiting distribution. For this, we need a doubly infinite version of the change process, which we denote by $\{C_n\}_{n=-\infty}^{\infty}$. For stationarity to hold, [23] and (2.2) show that X_0 must be formed from all past C_ns in the manner

$$X_0 = \sup(0, C_0, C_0 + C_{-1}, \ldots, C_0 + C_{-1} + \cdots + C_{-n}, \ldots). \tag{2.3}$$

Since X_0 is a function of $C_0, C_{-1}, \ldots$, we write the "causal" measurable function in (2.3) as $X_0 =$

$H_0(C_0, C_{-1}, \ldots)$. Likewise, $X_1 = \max(X_0 + C_1, 0)$ is a causal measurable function of $C_1, C_0, \ldots$; viz., $X_1 = h(C_1, C_0, \ldots)$. Generalizing this gives $X_k = H_k(C_k, C_{k-1}, \ldots)$ for a measurable function H_k.

For a general $k \geq 1$ and integer $h > 0$, we have

$$\begin{aligned}(X_0, X_1, \ldots, X_k) &= (H_0(C_0, C_{-1}, \ldots), h(C_1, C_0, \ldots), \ldots, H_k(C_k, C_{k-1}, \ldots)) \\ &\stackrel{\mathcal{D}}{=} (H_0(C_h, C_{h-1}, \ldots), h(C_{h+1}, C_h, \ldots), \\ &\quad \ldots, H_k(C_{k+h}, C_{k+h-1}, \ldots)) \\ &= (X_h, X_{h+1}, \ldots, X_{h+k}),\end{aligned}$$

where $\stackrel{\mathcal{D}}{=}$ indicates equality in distribution, and the equality in distribution conclusion follows by shifting the strictly stationary path of $\{C_t\}$ by h units.

This shows that when X_0 is in its stationary state as generated by the infinite history of the change process, $\{X_n\}_{n=0}^{\infty}$ is a strictly stationary process. A caveat here: one cannot take X_0 to be independent of $\{C_n\}_{n=-\infty}^{0}$ and obtain a strictly stationary $\{X_n\}_{n=0}^{\infty}$.

2.2 Markov Properties

This section studies the Markov structure of the Lindley random walk process $\{X_n\}$ with correlated changes. Clarifying our notion of Markov, a process $\{U_n\}_{n=0}^{\infty}$ is said to be Markov of order p if

$$P(U_n \leq x \mid U_{n-1}, \ldots, U_{n-p}, \ldots, U_0) = P(U_n \leq x \mid U_{n-1}, \ldots, U_{n-p})$$

for all real x, $n \geq 1$, and $U_0, \ldots, U_{n-1}$.

Before continuing, we need to introduce a class of strictly stationary time series models for $\{C_n\}$. The most widely used time series model class is the autoregressive moving-average (ARMA) models. ARMA models with an autoregressive order $p \geq 0$ and a moving-average order $q \geq 0$ are the unique (in mean square) solutions to the difference equation

$$C_n - \mu = \phi_1(C_{n-1} - \mu) + \ldots + \phi_p(C_{n-p} - \mu) + \epsilon_n + \theta_1\epsilon_{n-1} + \ldots + \theta_q\epsilon_{n-q}. \tag{2.4}$$

Here, $\{\epsilon_n\}$ is a sequence of IID random variables with zero mean and variance $\sigma^2 > 0$, $\phi_1, \ldots, \phi_p$ are

the p autoregressive coefficients, and $\theta_1, \ldots, \theta_q$ are the q moving-average coefficients. It is important that $\{\epsilon_n\}$ be IID and not simply uncorrelated noise; this will be needed in our Markov structure arguments below. We also assume a causal ARMA model. This stipulation requires that all roots of the autoregressive polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ lie outside the complex unit circle. For causal models, C_n can be expressed in terms of the current and past ϵ_ns only, namely $\epsilon_n, \epsilon_{n-1}, \ldots$.

For the ARMA(p, q) model in (2.4), $E[C_n] \equiv \mu$ for all n and $\{C_n\}$ is strictly stationary. For a stable model, we need $\mu < 0$ as shown in the last section (this is henceforth assumed). Computation of the autocovariances $\gamma(h) := \mathrm{Cov}(C_n, C_{n+h})$ proceeds from many classic algorithms. For more on this and other properties of ARMA series, see [3].

Many practitioners nowadays focus on autoregressions (AR) only due to their parsimonious and flexible structure and forecasting ease; that is, $\theta_1 = \cdots = \theta_q = 0$. It is easy to see that an AR(p) series is a Markov chain of order p when $\{\epsilon_n\}$ is IID; indeed, (2.4) explicitly writes X_n is a function of the p past series values and an independent noise that does not depend on past process values.

Perhaps the most commonly used time series model, and one that is used below, is a causal first order autoregression (AR(1)). This model obeys

$$C_n - \mu = \phi(C_{n-1} - \mu) + \epsilon_n, \tag{2.5}$$

where $\mid \phi \mid < 1$ is imposed for causality.

A more general class of strictly stationary change processes assumes the causal linear process structure

$$C_n = \mu + \sum_{k=0}^{\infty} \psi_k \epsilon_{n-k},$$

where the deterministic weight sequence $\{\psi_k\}_{k=0}^{\infty}$ satisfies $\sum_{k=0}^{\infty} \mid \psi_k \mid < \infty$. Reference [3] shows that any causal ARMA sequence with IID innovations has this representation and discusses the related Wold decomposition for stationary time series. Recursing (1.36) yields

$$X_n = \max\left(X_{n-L} + C_{n-L+1} + \cdots + C_n; C_{n-L+2} + \cdots + C_n; \ldots; C_n; 0\right), \tag{2.6}$$

which shows how X_n depends on X_{n-L} for any $L \geq 1$. Hence, in general, X_n and X_{n+h} will be dependent for all lags h, even when $\{C_n\}$ is m-dependent (say a moving-average of order m). Indeed, (2.6) implies that X_n and X_{n-L} will be dependent in general for any $L \geq 1$. Note however that this

dependence does not necessarily imply the absence of a Markov structure. For example, consider the causal AR(1) process described above. Recursing (2.5) gives

$$C_n - \mu = \phi^L(C_{n-L} - \mu) + \phi^{L-1}\epsilon_{n-L+1} + \cdots + \phi\epsilon_{n-1} + \epsilon_n.$$

for any $L \geq 1$. Thus, C_n and C_{n-L} are dependent; however, this process is known to be first order Markov [3].

Our first result shows that $\{X_n\}$ is not Markov of any order unless $\{C_n\}$ has additional structure. This corrects a mistaken claim in [26].

Lemma 2.2.1. *The general Lindley walk $\{X_n\}_{n=0}^{\infty}$ is not Markov of any order.*

Proof. To construct a counterexample, let $\{C_n\}$ be a general stationary Gaussian process. We consider the stable case where $\mu_C < 0$ and let $x_0, \ldots, x_{t-p-1}$ be arbitrary strictly positive feasible values for the process. Consider the conditional probability

$$P[X_t \leq y \mid X_{t-1} = 0, \ldots, X_{t-p} = 0; X_{t-p-1} = x_{t-p-1}, \ldots, X_0 = x_0] \tag{2.7}$$

(this takes $x_{t-p} = \cdots = x_{t-1} = 0$).

We now write this probability strictly in terms of the $\{C_n\}$ process. By (1.36) and the fact that $X_{t-1} = 0$, $\{X_t \leq y\} = \{C_t \leq y\}$ for all $y \geq 0$. Also, the event $[X_{t-1} = 0, \ldots, X_{t-p} = 0; X_{t-p-1} = x_{t-p-1}, \ldots, X_0 = x_0]$ can be written in terms of the C_ts via

$$\left[\cap_{i=1}^{t-p-1} C_i = x_i - x_{i-1} \cap C_{t-p} \leq -x_{t-p-1} \cap_{i=1}^{p-1} C_{t-p+i} \leq 0\right] \tag{2.8}$$

(more is said about this in the future sections). It follows that the probability in (2.7) equals

$$P\left[C_t \leq y \mid \cap_{i=1}^{t-p-1} C_i = x_i - x_{i-1} \cap C_{t-p} \leq -x_{t-p-1} \cap_{i=1}^{p-1} C_{t-p+i} \leq 0\right]. \tag{2.9}$$

For a general Gaussian stationary process, properties of multivariate normal densities show that the conditional probability in (2.9) will depend on x_{t-p-1} (and the previous x_ts too). As such, $\{X_t\}$ cannot be pth order Markov.

Even in cases where $\{C_t\}$ is first order Markov — say an AR(1) series — the conditional probability in (2.9) can be shown to depend on x_{t-p-1} when $\phi \neq 0$ due to the fact that we are

conditioning on some non-singleton sets in (2.8) (this takes additional work to see, which we do not pursue here). □

Our next result establishes the Markov structure of the bivariate process $\{(X_n, C_n)\}$ when $\{C_n\}$ is a pth order Markov chain.

Proposition 2.2.2. *The process $\{(X_n, C_n)\}$ is a pth order Markov chain when $\{C_n\}$ is a pth order Markov chain (such as the above AR(p) series).*

Proof. For clarity, we first provide the argument for the case where $p = 1$. For this p, (1.36) gives

$$X_{n+1} = \max(X_n + C_{n+1}, 0) := h(X_n, C_{n+1})$$

for the measurable function $h : \mathbb{R}^2 \to \mathbb{R}$ defined by $h(x, y) = \max(x + y, 0)$. From this and the first order Markov property of $\{C_n\}$, the joint dynamics of (X_{n+1}, C_{n+1}), given the entire history $C_1, \ldots, C_n$ and $X_0, \ldots, X_n$, are described solely by X_n and C_n — we do not need $C_1, \ldots, C_{n-1}$ or $X_0, X_1, \ldots, X_{n-1}$. The conclusion now follows.

When $p > 1$, merely extend the above logic by recursing (1.36) p times to get

$$X_{n+1} = h(X_{n-p+1}, C_{n-p+2}, \ldots, C_n, C_{n+1})$$

for a measurable function $h : \mathbb{R}^{p+1} \to \mathbb{R}$ (the form of h is not important, but one may wish to compare to (2.6) and argue as above. While here, we note that given that past p X_ns and C_ns, X_{n+1} can be written in a form that does not involve $X_{n-p+2}, \ldots, X_n$, but rather only X_{n-p+1} and $C_{n-p+2}, \ldots, C_{n+1}$. □

The above arguments also prove the following result

Proposition 2.2.3. *The $p+1$ dimensional process $\{(X_n, C_n, C_{n-1}, \ldots, C_{n+1-p})\}_{n=p}^{\infty}$ is a first order Markov chain whenever $\{C_n\}$ is a causal AR(p) series.*

2.3 The Likelihood for AR(p) Changes

The likelihood of a Lindley walk with IID $\{C_n\}$ was studied in [20], with other works since appearing [1]. Our goal in this section is two-fold. First, we clarify the support set of the distribution

of $\mathbf{X} = (X_0, \ldots, X_N)'$ and develop the likelihood function in terms of the distribution of the change process $(C_1, \ldots, C_N)'$ and the initial value X_0 (see (2.11)). In general, X_n has a point mass at zero and a possible density over $(0, \infty)$ for each fixed n. Secondly, we derive the likelihood function of the walk when $\{C_n\}$ is a pth order causal autoregression. The derivation makes use of the Markov properties in Section 2.2 and reduces the integration dimension in the general likelihood function in (2.11). While the obtained structure has some unwieldy multivariate integrals, approximation methods for the likelihood are studied in Section 2.4.

We now study the support set of $\mathbf{X}$. To avoid trite work, we assume that $\mathbf{C} = (\mathbf{C_1}, \ldots, \mathbf{C_N})'$ has the joint probability density function (PDF) $f_{\mathbf{C}}(\mathbf{c})$, where $\mathbf{c} = (\mathbf{c_1}, \ldots, \mathbf{c_N})' \in \mathbb{R}^{\mathbf{N}}$. The initial starting level X_0 is non-negative with cumulative distribution function $F_{X_0}(x)$; this distribution has a point mass at zero ($F_{X_0}(0) > 0$) and a density on $(0, \infty)$. For a strictly stationary $\{X_n\}$, X_0 must be a function of the past changes $C_0, C_{-1}, \ldots$, as is quantified in (2.3). The random vector $\mathbf{X}$ has a distribution that is a mixture of densities and mass functions on different domain regions. To quantify these regions, consider a partition of $\mathbb{R}_+^{N+1}$ defined as follows: let $I \subset \mathcal{I} \equiv \{0, 1, 2, \ldots, N\}$, and define a set associated with I via

$$B_I = \{\mathbf{y} \in \mathbb{R}_+^{N+1} : y_n > 0 \ \ \forall n \in I, \ \text{ and } \ y_n = 0 \ \forall n \notin I\}.$$

Here, I contains all indices with positive components. In particular, $B_{\mathcal{I}}$ is the interior of $\mathbb{R}_+^{N+1}$ and $B_\emptyset = \{\mathbf{0}\}$. Note that the $B_I, I \subset \mathcal{I}$, are disjoint and $\cup_{I \subset \mathcal{I}} B_I = \mathbb{R}_+^{N+1}$; thus, $\{B_I; I \subset \mathcal{I}\}$ partitions $\mathbb{R}_+^{N+1}$.

We note that the joint distribution constructed below is the likelihood of $\mathbf{X}$ as derived in Equation (3) of [11]. Now, consider an $I \subset \mathcal{I}$ that contains a non-empty sequence of indices $i_1, \ldots, i_k$ satisfying $0 \le i_1 < i_2 < \cdots < i_k \le N$. For $\vec{x} \in B_I$, the cumulative distribution function $F_{\mathbf{X}}(\vec{x}) = P[\cap_{i=0}^N X_i \le x_i]$ is differentiable in the variables $x_i, i \in I$, and we write

$$L_{\mathbf{X}}(\vec{x}) = \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} F_{\mathbf{X}}(\vec{x}) \tag{2.10}$$

for this density.

Define $L_{\mathbf{X}}(\mathbf{0}) = F_{\mathbf{X}}(\mathbf{0}) = P[X_0 = 0, \ldots, X_N = 0]$. For each $B \subset B_I$, the positive components in B can be obtained using the map $\mathcal{M}(B) = \{(x_{i_1}, \ldots, x_{i_k}) : \vec{x} \in B\}$. We regard $L_{\mathbf{X}}(\vec{x})$ as a

function of the positive component of $\vec{x}$ and write $L_{\mathbf{X}}(\vec{x}) = L_{\mathbf{X}}(x_{i_1}, \ldots, x_{i_k})$.

Now consider a probability measure μ_I, defined on $\mathbb{R}_+^{N+1}$, such that for $A \in \mathcal{B}(\mathbb{R}_+^{N+1})$,

$$\mu_I(A) = \int_{\mathcal{M}(A \cap B_I)} L_{\mathbf{X}}(\mathbf{x}) \lambda(d\mathbf{x}),$$

where λ is the Lebesgue measure on $\mathcal{M}(B_I)$. If $I = \emptyset$, λ reduces to a discrete Dirac measure. We observe that $\mu_I, I \subset \mathcal{I}$ are mutually singular. The distribution of $\mathbf{X}$ can thus be characterized by

$$\mathcal{L}(\mathbf{x}) = \sum_{I \subset \mathcal{I}} 1_{\{\mathbf{x} \in \mathbf{B_I}\}} L_{\mathbf{X}}(\mathbf{x});$$

this is what we consider as our likelihood. Because the data $\mathbf{X}$ are fixed when optimizing over the parameters $\mathbf{\Theta}$ in a likelihood, we also write $\mathcal{L}(\mathbf{\Theta})$ or $\mathcal{L}(\mathbf{\Theta} \mid \mathbf{X})$ for our likelihood. Other variants of notation are used below in obvious manners.

We now compute the likelihood of $\mathbf{X}$ using $f_{X_0,\mathbf{C}}(x, \mathbf{c}), \mathbf{x} \geq \mathbf{0}, \mathbf{c} \in \mathbb{R}^{\mathbf{N}}$ – the joint distribution of X_0 and $\mathbf{C}$ (when $x > 0$, $f_{X_0,\mathbf{C}}(x, \mathbf{c})$ represents the joint density function, and when $x = 0$, $f_{X_0,\mathbf{C}}(0, \mathbf{c}) = \partial^{\mathbf{N}} \mathbf{F}_{\mathbf{X_0},\mathbf{C}}(\mathbf{0}, \mathbf{c}) / \partial \mathbf{c_1} \cdots \partial \mathbf{c_N}$). We are given the observations $\mathbf{X} = (\mathbf{X_0}, \mathbf{X_1}, \ldots, \mathbf{X_N})' \in \mathbf{B_I}$. Recall that $I = \{i_1, \ldots, i_k\}$ such that $X_m > 0$ for $m \in I$ and $X_n = 0$ for $n \notin I$. Write $I^c \equiv \mathcal{I} \backslash I = \{j_1, \ldots, j_{N-k}\}$. The CDF of $\mathbf{X}$ at $\vec{x}$ is

$$F_{\mathbf{X}}(\vec{x}) = P(X_0 \leq x_0, X_m \leq x_m, m \in I, X_n = 0, n \in I^c).$$

Without loss of generality, assume $x_0 > 0$. From (2.10), the likelihood function can be derived as the following limit:

$$\mathcal{L}(\mathbf{\Theta} \mid \mathbf{X} = \tilde{\mathbf{x}}) = \lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P(X_m \in (x_m - h/2, x_m + h/2), m \in I \cup \{0\},$$
$$X_n = 0, n \in I^c),$$

where $\alpha = |I \cup \{0\}|$, and h is assumed to be smaller than $2x_m$ for all m so that each interval $(x_m - h/2, x_m + h/2)$ is nonempty. From (1.36), for $m \in I$, $x_m > 0$, implying that $C_m = X_m - X_{m-1}$;

for $n \in I^c$, $x_n = 0$, implying that $C_n \leq -X_{n-1}$. This implies that

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\Theta} \mid \mathbf{X} = \tilde{\mathbf{x}}) = \\
&\lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P\Big(X_m \in (x_m - h/2, x_m + h/2), C_m = X_m - X_{m-1}, m \in I, \\
&\qquad\qquad X_n = 0, C_n \leq -X_{n-1}, n \in I^c, X_0 \in (x_0 - h/2, x_0 + h/2) \Big) = \\
&\lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P\Big(X_m \in (x_m - h/2, x_m + h/2), C_m = X_m - X_{m-1}, m \in I, \\
&\qquad\qquad C_n \leq -x_{n-1}, n \in I^c, X_0 \in (x_0 - h/2, x_0 + h/2) \Big).
\end{aligned}$$

Using the change of variable formula for $C_m = X_m - X_{m-1}, m \in I$, we have the likelihood function

$$\mathcal{L}(\boldsymbol{\Theta} \mid \mathbf{X}) = \int_{-\infty}^{-X_{j_1 - 1}} \cdots \int_{-\infty}^{-X_{j_{N-k}-1}} f_{X_0, \mathbf{C}}(X_0, X_m - X_{m-1}, m \in I, \mathbf{c}) d\mathbf{c}, \tag{2.11}$$

where the integral is over the set $\{\mathbf{c} \in \mathbb{R}^{N-k} : C_n \leq -X_{n-1}, n \in I^c\}$. If X_0 is independent of $\mathbf{C}$, the integrand in (2.11) becomes $f_{X_0}(X_0) f_{\mathbf{C}}(X_0, X_m - X_{m-1}, m \in I, \mathbf{c})$. The formula in (2.11) involves high dimensional multiple integrals when $\mathbf{X}$ has many zeros even after employing the Markov properties of Section 2.2. If one is to assume a Gaussian changes process $\{C_n\}$, the high dimensional integrals are known to cause considerable difficulty [10].

We now move to deriving the likelihood when $\{C_n\}$ is an autoregressive process of order p satisfying the AR(p) recursion in (2.4) — the process need not be Gaussian. The Markov structure identified in Section 3 effectively reduces the integral dimension in (2.11). Our goal is to explicitly derive the likelihood in terms of its free parameters, which are $\boldsymbol{\Theta} = (\mu, \sigma^2, \phi_1, \ldots, \phi_p)'$.

Recall that if $X_n > 0$, the value of C_n can be recovered; namely, $C_n = X_n - X_{n-1}$; when $X_n = 0$, we only know that the value of C_n is less than or equal to $-X_{n-1}$, but we do not know its exact value. Previous authors [26, 34, 5] have viewed this problem as a censored time series issue. The censoring here is not simple; indeed, when $X_n = 0$, the values of C_n are censored depending on $-X_{n-1}$, which is not constant in time and also depends on the past history of the process.

Our derivation partitions the series into segments where C_n is "*recoverable*" or not. We take X_n as observed for all $n \in \{0, \ldots, N\}$. While the derivation is somewhat tedious, this is expected given the difficulties encountered in likelihood evaluation in [26, 34] and [5]. Define the first *ariser*

time as

$$\kappa_1 := \min_{n \geq p} \{n : X_n > 0, \ldots, X_{n-p+1} > 0\},$$

which is the first time that p consecutive changes are recoverable. Next, define the first *plunger* time as

$$\tau_1 := \min_{n > \kappa_1} \{n : X_n = 0\}.$$

For $i \geq 1$, define the successive ariser and plunger times as

$$\kappa_{i+1} := \min\{n > \tau_i : X_n > 0, \ldots, X_{n-p+1} > 0\}$$

and

$$\tau_{i+1} := \min\{n > \kappa_{i+1} : X_n = 0\}.$$

Cases where κ_i or τ_i do not occur are addressed below.

Let $K(N) := \max\{i : \kappa_i \leq N\}$. Then the i^{th} complete regime of the process contains the time points in $R_i := \{\kappa_i + 1, \ldots, \tau_i, \ldots, \kappa_{i+1}\}$ for $i = 1, \ldots, K(N) - 1$. As bookends, set $R_0 = \{1, \ldots, \kappa_1\}$ and $R_{K(N)} = \{\kappa_{K(N)} + 1, \ldots, N\}$ if $k(N) < N$, and set $R_{K(N)} = \emptyset$ otherwise. Note that R_0 does not contain X_0. It is not possible to recover C_0 since X_{-1} is unobserved. Likewise, if κ_i does not exist (occur), set $R_0 = \{1, \ldots, N\}$. This blocks the observations into distinct regimes via its ariser times.

We now need some notation. Let $A = \{n_1, n_2, \ldots, n_k\}$ for $n_1 < n_2 < \cdots < n_k$ denote k ordered index times in $\{1, \ldots, N\}$. Let $\mathbf{X}_A = (X_{n_1}, \ldots, X_{n_k})'$ denote a $k \times 1$ vector of the ordered X_{n_i} for $n_i \in A$. We use the notation $\mathbf{X}_{A-1} = (X_{n_1-1}, \ldots, X_{n_k-1})'$. For convenience, let $\mathbf{X_n} = (X_0, X_1, \ldots, X_n)'$. We use the same notation for the $\{C_n\}$ process: $\mathbf{C_A} = (C_{n_1}, \ldots, C_{n_k})'$ for $A = \{n_1, n_2, \ldots, n_k\}$ and $\mathbf{C_n} = (C_1, \ldots, C_n)'$. For realized values, lowercase notation is used, e.g., $\mathbf{x}_n = (x_0, x_1, \ldots, x_n)'$ and $\mathbf{c_n} = (c_1, \ldots, c_n)'$. For each $n \in \mathbb{N}$, let $n(p) = \{n - p + 1, \ldots, n\}$ denote the p consecutive time points ending at time n. For two random variables/vectors Y and Z, the conditional "density" of Y given $Z = z$ is denoted by $f_{Y|Z=z}(\cdot \mid Z = z)$.

At time $n = \kappa_i$, $C_{\kappa_i}, \ldots, C_{\kappa_i - p+1}$ are observed and

$$\mathbf{C}_{\kappa_i(p)} = \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1} = (X_{\kappa_i - p+1} - X_{\kappa_i - p}, \ldots, X_{\kappa_i} - X_{\kappa_i - 1})',$$

where $\kappa_i(p) = \{k_i - p + 1, \ldots, \kappa_i\}$. Define the set of *good times* $G_i \subset R_i$ during the i^{th} regime as

$$G_i = \{n : \kappa_i < n < \tau_i\} = \{k_i + 1, \ldots, \tau_i - 1\},$$

where for each $n \in G_i$, $X_n, \ldots, X_{n-p} > 0$, implying that $C_n, \ldots, C_{n-p}$ are all recoverable. We use the convention $G_i^c = \{n : n \in R_i \cap n \notin G_i\} = \{\tau_i, \ldots, \kappa_{i+1}\}$ for ease of exposition. That is, the complement of the good times of the ith regime only contains times in the ith regime.

Denote by $\mathcal{L}_{G_i|\kappa_i}(\cdot \mid \mathbf{X}_{\boldsymbol{\kappa_i}})$ the conditional distribution of $\mathbf{X}_{\mathbf{G_i}}$ given the past $\mathbf{X}_{\boldsymbol{\kappa_i}}$. When $G_i \neq \emptyset$, the change of variables formula and the pth order Markov property of $\{C_n\}$ give

$$\begin{aligned}
&\mathcal{L}_{G_i|\kappa_i}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i}) \qquad (2.12)\\
&= f_{\mathbf{X}_{G_i}|\mathbf{X}_{\kappa_i}}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i})\\
&= f_{\mathbf{X}_{G_i}|\mathbf{C}_{\kappa_i(p)},\mathbf{X}_{\kappa_i-p}}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}, \mathbf{X}_{\kappa_i-p})\\
&= f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)},\mathbf{X}_{\kappa_i-p}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}, \mathbf{X}_{\kappa_i-p})\\
&= f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}), \qquad (2.13)
\end{aligned}$$

where $f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\cdot \mid \cdot)$ is the conditional "density" of the changes during the good times G_i in R_i conditional on the p consecutive changes $\mathbf{C}_{\kappa_i(p)}$ ending at the ariser time κ_i prior to the start of R_i. One may use the fact that $\{C_n\}$ is p^{th} order Markov to arrive at

$$\begin{aligned}
&f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}) =\\
&\prod_{n \in G_i} f_{C_n|\mathbf{C}_{(n-1)(p)}}(X_n - X_{n-1} \mid \mathbf{X}_{(n-1)(p)} - \mathbf{X}_{(n-1)(p)-1}),
\end{aligned}$$

where $(n-1)(p) := \{n - p, \ldots, n - 1\}$. If $G_i = \emptyset$, the convention $f_{\mathbf{X}_{G_i}|\mathbf{X}_{\kappa_i}}(\cdot) = 1$ is assumed.

We will further decompose the times in G_i^c in the i^{th} regime into two sets of times. The first set collects times where the observations are positive, and the other one contains the times where the observations are zero. More precisely, define $\eta_i \subset G_i^c$ as $\eta_i = \{n \in G_i^c : x_n > 0\}$. Then η_i contains the times in regime R_i where the C_n are observed, but the previous p changes are not all observed. Likewise, define $\eta_i^c = \{n \in G_i^c : X_n = 0\}$ as those times in R_i where C_n is not observed and not all of the previous p changes are observed. Note that our definitions form a partition of each R_i: $R_i = G_i \cup \eta_i \cup \eta_i^c$.

To study the conditional distribution of $\mathbf{X}_{\mathbf{G_i^c}}$, we need to identify the zeros in the observations $\mathbf{X}_{\mathbf{G_i^c}} = (\mathbf{X}_{\tau_\mathbf{i}}, \ldots, \mathbf{X}_{\kappa_{\mathbf{i+1}}})'$. Let $z_1^i < \ldots < z_{L_i}^i$ denote the indices of those zeros. Clearly, $z_1^i = \tau_i$ and $z_{L_i}^i = \kappa_{i+1} - p$. Then $\eta_i = G_i^c / \{z_1^i, \ldots, z_{L_i}^i\}$ and $\eta_i^c = \{z_1^i, \ldots, z_{L_i}^i\}$. Let $\mathcal{L}_{G_i^c|(\tau_i-1)}(\cdot \mid \mathbf{X}_{\tau_\mathbf{i}-\mathbf{1}})$ denote the distribution of $\mathbf{X}_{\mathbf{G_i^c}}$ given $\mathbf{X}_{\tau_\mathbf{i}-\mathbf{1}}$. Similar to the analysis producing (2.11), and using the Markov property of $\{C_n\}$,

$$\begin{aligned}
&\mathcal{L}_{G_i^c|(\tau_i-1)}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1}) \\
&= \int_{-\infty}^{-X_{z_1^i-1}} \cdots \int_{-\infty}^{-X_{z_{L_i}^i-1}} f_{\mathbf{C}_{G_i^c}|\mathbf{C}_{(\tau_i-1)(p)}}(\mathbf{X}_{\eta_i} - \mathbf{X}_{\eta_i-1}, \mathbf{c} \mid \\
&\qquad\qquad \mathbf{X}_{(\tau_i-1)(p)} - \mathbf{X}_{(\tau_i-1)(p)-1}) d\mathbf{c}, \qquad (2.14)
\end{aligned}$$

where the integration domain is $\{\mathbf{c} \in \mathbb{R}^{L_i} : C_n \leq -X_{n-1}, n \in \eta_i^c\}$ and $(\tau_i - 1)(p) = \{\tau_i - p, \ldots, \tau_i - 1\}$. The conditional density in (2.14) can also be written as a product of conditional density $f_{C_n|\mathbf{C}_{(n-1)(p)}}(\cdot)$ for $n \in G_i^c$. We omit this detail. If $G_i^c = \emptyset$, then set $\mathcal{L}_{G_i^c|(\tau_i-1)}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1})$ to unity (this can only happen in R_0 or $R_{K(N)}$).

We now consider the 0^{th} startup regime. The distribution of X_0, which is the stationary distribution of the process, is absolutely continuous away from zero and has a point mass at zero. While the distribution of X_0 does not have an explicit form, its moment properties are studied in [2] and [32]. Until the first time n such that p consecutive C_ns are observed (which happens at the time $n = \kappa_1$), the joint distribution of $(X_0, X_1, X_2, \ldots, X_{\kappa_1})'$ depends on the joint distribution of $(X_0, C_1, \ldots, C_{\kappa_1})'$ according to (2.11). More precisely, let $\eta_0 = \{n \in R_0 : X_n > 0\}$ and $\eta_0^c := \{n \in R_0 : X_n = 0\} = \{z_1, \ldots, z_L\}$. Then from (2.11) the likelihood of $\mathbf{X}_{R_0}$ is

$$\mathcal{L}_{R_0}(\vec{x}_{R_0}) = \int_{-\infty}^{-X_{z_1-1}} \cdots \int_{-\infty}^{-X_{z_L-1}} f_{X_0, \mathbf{C}_{\kappa_\mathbf{1}}}(X_0, X_m - X_{m-1}, m \in \eta_0, \mathbf{c}) d\mathbf{c}, \qquad (2.15)$$

where $\mathbf{C}_{\kappa_1} = (C_1, \ldots, C_{\kappa_1})'$ and the integration domain is $\{\mathbf{c} \in \mathbb{R}^L : c_n \leq -X_{n-1}, n \in \eta_0^c\}$.

As a summary, combining (2.13), (2.14), and (2.15), our likelihood is

$$\mathcal{L}(\mathbf{\Theta} \mid \mathbf{X}) = \mathcal{L}_{R_0}(\mathbf{X}_{R_0}) \times \left(\prod_{i=1}^{K(N)} \mathcal{L}_{G_i|\kappa_i}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i}) \times \mathcal{L}_{G_i^c|\tau_i-1}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1}) \right). \qquad (2.16)$$

Our expression for the likelihood is inconvenient due to the integrals needed to evaluate the multivariate joint CDF's (as needed to numerically evaluate Equations (2.14) and (2.15) for

example). Because of this, our next section moves to a technique that efficiently simulates this likelihood to a degree where statistical inferences can be accurately made.

2.4 Particle Filtering Likelihood Evaluation

This section introduces particle filtering methods to approximate and optimize the storage model's likelihood. The exact likelihood in (2.16) is problematic to evaluate due to its multiple integrals. Here, we construct an approximation to it, viewing the problem as a censored time series issue.

2.4.1 Particle Filtering Methods

As noted in the last section, C_n can be recovered exactly as $X_n - X_{n-1}$ when $X_n > 0$. When $X_n = 0$, we know that $C_n \leq -X_{n-1}$, but we do not know C_n exactly. For such times n, define a censoring indicator $\delta_n = 1$; set $\delta_n = 0$ if C_n can be recovered exactly at time n. For convenience, our initial condition sets $C_1 = X_1$ and we work with the data $X_1, \ldots, X_N$.

For notation, let $\pi_1, \ldots, \pi_r$ denote the ordered times at which C_n is censored and $d_1, \ldots, d_s$ be the ordered times at which C_n can be exactly recovered. Obviously, $r + s = N$. Define the censored series as

$$C_n^* = \begin{cases} X_n - X_{n-1}, & \text{if } C_n \text{ is recoverable} \\ -X_{n-1}, & \text{if } C_n \text{ is not recoverable} \end{cases}.$$

The likelihood of $(C_1, \ldots, C_N)'$ can be written in terms of the uncensored C_ns as

$$\mathcal{L}(\boldsymbol{\Theta}) = \int_{\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}), i \in \{1,\ldots,r\}\}} \mathcal{N}_{\boldsymbol{\Theta}}(c_{1:N}) dc_{\pi_1} \ldots dc_{\pi_r}, \tag{2.17}$$

where we have taken $\{C_n\}$ to be a Gaussian process; this agrees with (2.11). Here, a joint Gaussian probability density for $(C_1, \ldots, C_N)'$, denoted by $\mathcal{N}_{\boldsymbol{\Theta}}(c_{1:N})$, is assumed. While other marginal distributions for $\{C_n\}$ are possible, a Gaussian setup is the usual assumption in time series and is computationally convenient. Our problem now reduces to evaluating (2.17).

Literature to evaluate (2.17) includes [26]. Below, a novel particle filtering method that exploits the autoregressive structure of the series will be devised. We begin with importance sampling,

observing that

$$\begin{aligned}\mathcal{L}(\mathbf{\Theta}) &= \int_{\{c_{\pi_i}\in(-\infty,c^*_{\pi_i}),i\in\{1,\ldots,r\}\}} \mathcal{N}_{\mathbf{\Theta}}\left(c_{1:N}\right) dc_{\pi_1}\ldots dc_{\pi_r} \\ &= \int_{\{c_{\pi_i}\in(-\infty,c^*_{\pi_i}),i\in\{1,\ldots,r\}\}} \frac{\mathcal{N}_{\mathbf{\Theta}}\left(c_{1:N}\right)}{q(c_{\pi_1},\ldots,c_{\pi_r})} q(c_{\pi_1},\ldots,c_{\pi_r}) dc_{\pi_1}\ldots dc_{\pi_r},\end{aligned}$$

where $q(\cdot)$ is any probability density function that we call a proposal density function and $c_{1:N} = (c_1,\ldots,c_N)'$. We want $q(\cdot)$ to be easy to sample from and to be supported over the set $\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}], i \in \{1,\ldots,r\}\}$.

Assume that M independent samples are drawn from $q(\cdot)$. Then a law of large numbers approximation of the likelihood in (2.17) is

$$\mathcal{L}(\mathbf{\Theta}) = E_q[W] \approx \frac{1}{M}\sum_{m=1}^{M} \frac{\mathcal{N}_{\mathbf{\Theta}}\left(c^{(m)}_{1:N}\right)}{q(c^{(m)}_{\pi_1},\ldots,c^{(m)}_{\pi_r})},$$

where $W = \mathcal{N}_{\mathbf{\Theta}}\left(C_{1:N}\right)/q(C_{\pi_1:\pi_r})$ is viewed as a "weight". The subscript of q on E implies that the expectation is taken relative to the distribution q. In our notation, superscripts of (m) refer to the mth generated sample (of M total).

We need to develop a "nice" $q(\cdot)$ to facilitate our sampling procedure. To do this, we consider an AR(1) scheme to illustrate the ideas; this is easily extendable to AR(p) settings. In the AR(1) case, $\mathbf{\Theta}$ contains the three parameters μ, ϕ, and σ^2. The proposal density we use is the conditional probability density of the uncensored data $C_{\pi_1},\ldots,C_{\pi_r}$ given the censored values $C^*_{\pi_1},\ldots,C^*_{\pi_r}$:

$$q(c_{\pi_1},\ldots,c_{\pi_r}) = p(c_{\pi_1},\ldots,c_{\pi_r} \mid c^*_{\pi_1},\ldots,c^*_{\pi_r}),$$

where $p(\cdot \mid \cdot)$ is used as notation for a generic conditional probability density function. Assuming $\pi_1 \neq 1$, the first order Markov property for AR(1) series provides

$$p(c_{\pi_1},\ldots,c_{\pi_r} \mid c^*_{\pi_1},\ldots,c^*_{\pi_r}) = \prod_{i=1}^{r} p(c_{\pi_i} \mid c_{\pi_i-1}, c^*_{\pi_i}). \tag{2.18}$$

To see (2.18), note that c_{π_i-1} and the parameters in $\mathbf{\Theta}$ determine the normal distribution's mean and standard deviation of c_{π_i}, and $c^*_{\pi_i}$ indicates the upper bound of the truncation. Hence, $p(c_{\pi_i} \mid c_{\pi_i-1}, c^*_{\pi_i})$ is simply the truncated normal density with support on $(-\infty, c^*_{\pi_i})$, having a mean and

variance that are the one-step-ahead prediction of C_{π_i} from C_{π_i-1}. To further see this, note that

$$p(c_{\pi_i} \mid c_{\pi_i-1}, c^*_{\pi_i}) = \frac{\varphi\left(c_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}{\Phi\left(c^*_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right) - \Phi(-\infty)} = \frac{\varphi\left(c_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}{\Phi\left(c^*_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}.$$

where $\hat{m}_{\pi_i} = \phi C_{\pi_i-1}$ and $\hat{r}_{\pi_i} = \sigma^2$ are the one-step-ahead predictions and variances of C_{π_i} from $C_{\pi_{i-1}}$. Here, we have used φ and Φ as notation for the standard normal density and cumulative distribution functions.

A complication here is that $C_{\pi_{i-1}}$ may or may not be censored. If $C_{\pi_{i-1}}$ is not censored, we use the observation $C_{\pi_{i-1}}$; otherwise, we use the generated value of $C_{\pi_{i-1}}$, which is always available from the sampling adopted.

Hence, the proposal distribution in (2.18) is relatively easy to sample from, with weight

$$\begin{aligned} W^{(m)} &= \frac{\mathcal{N}_{\boldsymbol{\Theta}}\left(c^{(m)}_{1:N}\right)}{q\left(c^{(m)}_{\pi_1:\pi_r}\right)} = \frac{\prod_{j=1}^{N} p\left(c^{(m)}_{j} \mid c^{(m)}_{j-1}\right)}{\prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}, c_{\pi^*_i}\right)} \\ &= \frac{\prod_{j=1}^{s} p(c_{d_j} \mid c_{d_j-1}) \prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}\right)}{\prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}, c_{\pi^*_i}\right)}. \end{aligned}$$

Equation (2.18) and $p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}\right) = \varphi\left(c^{(m)}_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right)$ give

$$\frac{p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}\right)}{p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i-1}, c_{\pi^*_i}\right)} = \Phi\left(c^*_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right).$$

Therefore, our form for the weight is

$$W^{(m)} = \prod_{j=1}^{s} \varphi\left(c_{d_j} \mid \hat{m}^{(m)}_{d_j}, \hat{r}^{(m)}_{d_j}\right) \times \prod_{i=1}^{r} \Phi\left(c^*_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right).$$

Summarizing, our complete algorithm for the AR(1) case is as follows:

1. if $\delta_i = 1$, C_1 is uncensored and $W_1 = \varphi(C_1 \mid -\mu, \sigma^2/(1-\phi^2))$; if $\delta_i = 0$, C_1 is censored and $W_1 = \Phi(C_1 \mid -\mu, \sigma^2/(1-\phi^2))$, and C_1 is sampled from truncated normal density $\mathcal{N}\left(\hat{m}_{\pi_i}, \hat{r}_{\pi_i}; -\infty, 0\right)$.

After step 1, repeat steps 2 and 3 until $i = N$:

2. if $\delta_i = 1$, X_i is uncensored and update via

$$W_i = W_{i-1} \times \varphi(c_i \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}),$$

if $\delta_i = 0$, X_i is censored and update via

$$W_i = W_{i-1} \times \Phi(c_i^* \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}),$$

3. if $\delta_i = 1$, do nothing; if $\delta_i = 0$, sample C_i from the truncated normal distribution

$$\mathcal{N}\left(\hat{m}_{\pi_i}, \hat{r}_{\pi_i}; -\infty, c_i^*\right)$$

4. Finally, record W_N.

The above process is repeated M times, where M is so large that law of large number approximations are good. Generally, the larger M is, the better the approximation will be. The final approximated likelihood is

$$\mathcal{L}(\boldsymbol{\Theta}) \approx \frac{1}{M} \sum_{m=1}^{M} W_N^{(m)}.$$

Before closing, we comment on a naive way to simulate the likelihood with a Gaussian $\{C_n\}$. Another way to evaluate (2.17) decomposes the likelihood as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\Theta}) = \int_{\{c_{\pi_i} \in (-\infty, c_{\pi_i}^*)\}} & \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{\pi_1}, \cdots, c_{\pi_p} \mid c_{d_1}, \cdots, c_{d_p}\right) \\
& \times \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{d_1}, \cdots, c_{d_p}\right) dc_{\pi_1} \cdots dc_{\pi_p} \\
= \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{d_1}, \cdots, c_{d_p}\right) & \int_{\{c_{\pi_i} \in (-\infty, c_{\pi_i}^*)\}} \mathcal{N}_{\boldsymbol{\Theta}}\big(c_{\pi_1}, \cdots, c_{\pi_p} \mid \\
& \qquad c_{d_1}, \cdots, c_{d_p}\big) dc_{\pi_1} \cdots dc_{\pi_p}
\end{aligned}$$

and uses the explicit form of the conditional multivariate normal density in [31] to sample the censored values conditional on the uncensored values. This is more computationally expensive than the proposed particle filtering method because drawing from the conditional normal distribution may require inverting a large covariance matrix that is of dimension equal to the number of censored values. When μ is highly negative and N is large, many data points will be censored and this

dimension may be large.

2.4.2 A Simulation Study

This subsection presents a simulation study that evaluates the performance of our particle filtering estimation methods in the last subsection. The $\mathbb{R}$ code and seed used to generate the data used in our analysis are available on request. We consider the case of a Gaussian AR(1) $\{C_n\}_{n=1}^N$. The parameters in this setup are μ, ϕ, and σ^2. Simulations for higher order autoregressions have the same good performance as the AR(1) simulations, but are not shown here for brevity's sake.

The mean μ is taken as negative to ensure that $\{X_n\}$ is stable. The more negative μ is, the more frequently X_n will be zero. The AR(1) correlation parameter ϕ satisfies $\mid \phi \mid < 1$, which is needed for a causal $\{C_n\}$. We will examine $\phi \in \{-0.5, -0.25, 0, 0.25, 0.5, 0.75\}$, although negative ϕ do not arise in practice as much as positive ϕ. In all simulations, σ^2 is taken as unity.

Each simulated series uses $M = 10,000$ independent particles. The series lengths $n = 100, 250$, and 500 were studied. Common random number (CRN) techniques were used to the ensure that the likelihood is relatively smooth with respect to its parameters. This is an essential step with particle filtering methods — see [24] and [12] for more on CRNs. Finally, the popular quasi-Newton method L-BFGS-B is implemented to optimize the likelihoods. The true model parameters were used as initial guesses in our optimizations. It takes, on average, 15s, 45s, and 90s in the coding language R on a Macbook Pro computer to complete an analysis for one simulated series of length $n = 100, 250$, and 500, respectively.

Figure 2.1 shows boxplots of parameter estimators aggregated from 200 independent series. The sample means of the estimators are all close to their true values. Some minor bias is present in some cases, but this generally decays with increasing sample size. We remind the reader that likelihood estimation of AR(1) parameters in uncensored settings is also slightly biased (see [30] for a quantification).

For standard errors of the estimators, Table 2.1 reports two values: 1) the sample standard deviations of the parameter estimators over the 200 runs (denominator of 199), and 2) the average (over the 200 runs) of standard errors obtained by inverting the Hessian matrix at the maximum likelihood estimate for each run (denominator of 200). These two standard errors are close to one another, providing comfortable agreement.

Overall, the performance of the particle filtering estimation for AR(1) series is stellar. One

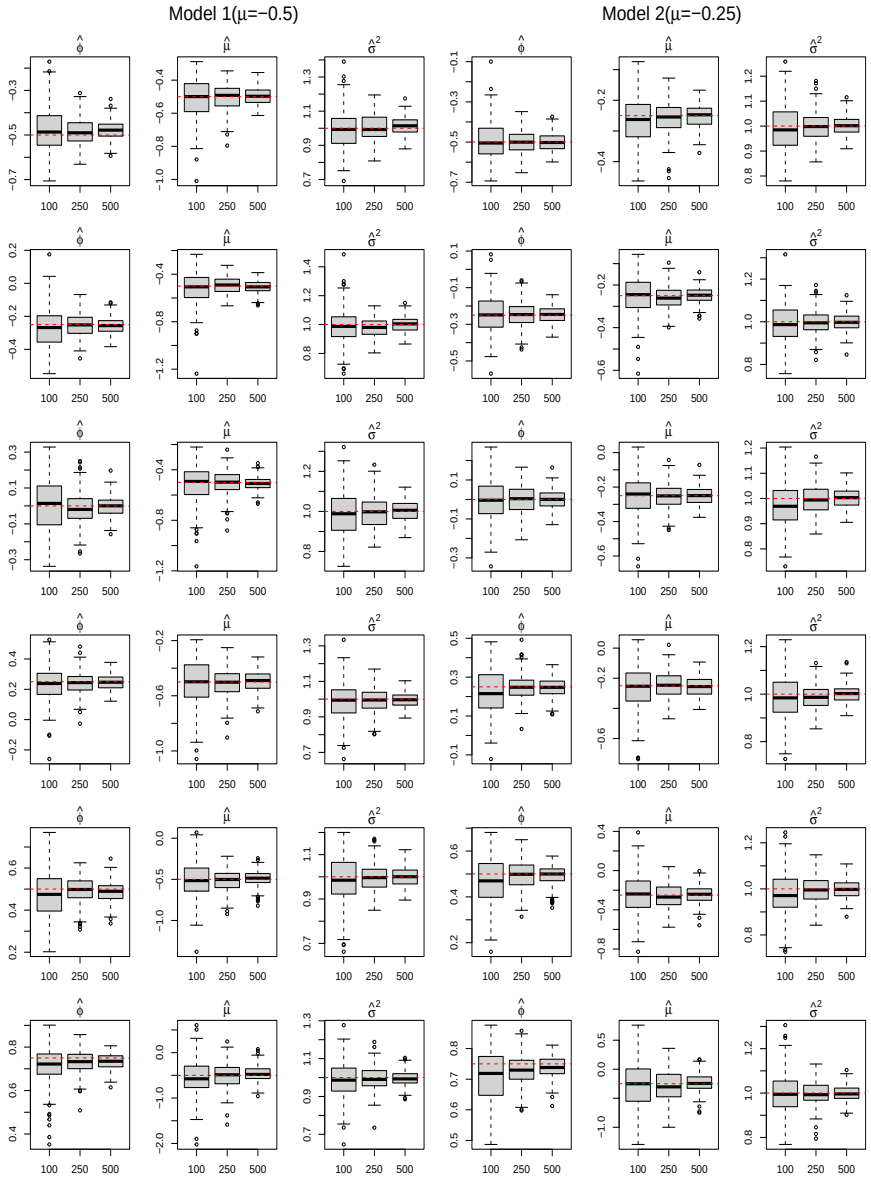

Figure 2.1: Boxplots of parameter estimators for the Lindley walk with an AR(1) $\{C_n\}$. The dashed lines demarcate true parameter values. All estimators appear roughly unbiased, with any bias decaying with increasing sample size. Columns 1-3 show results for $\mu = -0.5$, columns 4-6 show results when $\mu = -0.25$.

can even get an accurate standard error from one realization of the series by inverting the Hessian matrix at the likelihood estimators. In many particle filtering applications, "particle degeneration" occurs for larger n and results can degrade for these sample sizes. Methods to correct for particle degeneration are discussed in [7, 25, 27]; these do not appear needed here. This shows that likelihood

inference, the gold standard for statistical estimation, can be conducted for storage models.

			Model 1 ($\mu = -0.5$)			Model 2 ($\mu = -0.25$)		
ϕ	n		$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}^2$
-0.5	100	mean	-0.4799	-0.5153	0.9894	-0.4936	-0.2644	0.9886
		SD	0.0999	0.1223	0.1154	0.0944	0.0782	0.0912
		$\hat{E}(I'(\theta)^2)$	0.1028	0.1111	0.1166	0.0935	0.0812	0.0912
	250	mean	-0.4852	-0.5043	1.0041	-0.5000	-0.2563	0.9977
		SD	0.0606	0.0774	0.0750	0.0578	0.0526	0.0564
		$\hat{E}(I'(\theta)^2)$	0.0629	0.0685	0.0726	0.0585	0.0505	0.0571
	500	mean	-0.4786	-0.4982	1.0120	-0.4996	-0.2523	1.0023
		SD	0.0411	0.0498	0.0534	0.0419	0.0372	0.0387
		$\hat{E}(I'(\theta)^2)$	0.0433	0.0482	0.0512	0.0410	0.0355	0.0402
-0.25	100	mean	-0.2672	-0.5191	0.9868	-0.2474	-0.2480	0.9926
		SD	0.1243	0.1319	0.1237	0.1080	0.0915	0.0886
		$\hat{E}(I'(\theta)^2)$	0.1203	0.1207	0.1172	0.1071	0.0925	0.0897
	250	mean	-0.2536	-0.4934	0.9801	-0.2455	-0.2621	0.9964
		SD	0.0708	0.0721	0.0666	0.0702	0.0529	0.0584
		$\hat{E}(I'(\theta)^2)$	0.0755	0.0728	0.0715	0.0681	0.0585	0.0571
	500	mean	-0.2559	-0.5078	1.0025	-0.2487	-0.2502	0.9981
		SD	0.0489	0.0523	0.0532	0.0468	0.0372	0.0424
		$\hat{E}(I'(\theta)^2)$	0.0530	0.0525	0.0517	0.0478	0.0408	0.0399
0	100	mean	0.0022	-0.5182	0.9905	-0.0041	-0.2493	0.9706
		SD	0.1374	0.1456	0.1108	0.1072	0.1140	0.0868
		$\hat{E}(I'(\theta)^2)$	0.1282	0.1374	0.1126	0.1133	0.1092	0.0870
	250	mean	-0.0140	-0.5043	0.9944	-0.0004	-0.2543	0.9959
		SD	0.0890	0.0920	0.0762	0.0737	0.0721	0.0566
		$\hat{E}(I'(\theta)^2)$	0.0809	0.0833	0.0703	0.0713	0.0698	0.0557
	500	mean	-0.0012	-0.5090	1.0019	0.0000	-0.2502	1.0017
		SD	0.0532	0.0520	0.0491	0.0494	0.0543	0.0384
		$\hat{E}(I'(\theta)^2)$	0.0574	0.0594	0.0499	0.0503	0.0493	0.0393
0.25	100	mean	0.2405	-0.4988	0.9892	0.2220	-0.2609	0.9860
		SD	0.1194	0.1639	0.1026	0.1113	0.1426	0.0927
		$\hat{E}(I'(\theta)^2)$	0.1239	0.1626	0.1047	0.1113	0.1405	0.0869
	250	mean	0.2416	-0.5071	0.9921	0.2483	-0.2450	0.9875
		SD	0.0774	0.1032	0.0712	0.0637	0.0881	0.0520
		$\hat{E}(I'(\theta)^2)$	0.0787	0.1011	0.0660	0.0691	0.0891	0.0534
	500	mean	0.2454	-0.4962	0.9944	0.2464	-0.2566	1.0010
		SD	0.0512	0.0717	0.0461	0.0512	0.0672	0.0377
		$\hat{E}(I'(\theta)^2)$	0.0549	0.0707	0.0462	0.0489	0.0636	0.0383
0.5	100	mean	0.4702	-0.5056	0.9867	0.4689	-0.2466	0.9793
		SD	0.1064	0.2134	0.1019	0.1020	0.1932	0.0937
		$\hat{E}(I'(\theta)^2)$	0.1099	0.2194	0.0980	0.1001	0.2010	0.0837
	250	mean	0.4941	-0.5148	0.9955	0.4947	-0.2683	0.9942
		SD	0.0639	0.1289	0.0606	0.0631	0.1250	0.0537
		$\hat{E}(I'(\theta)^2)$	0.0679	0.1403	0.0614	0.0616	0.1317	0.0527
	500	mean	0.4863	-0.4908	1.0009	0.4948	-0.2444	0.9968
		SD	0.0468	0.0977	0.0443	0.0440	0.0906	0.0380
		$\hat{E}(I'(\theta)^2)$	0.0461	0.0962	0.0428	0.0429	0.0920	0.0366
0.75	100	mean	0.7119	-0.5585	0.9867	0.7053	-0.2611	0.9992
		SD	0.0885	0.3957	0.0992	0.0897	0.4161	0.0940
		$\hat{E}(I'(\theta)^2)$	0.0840	0.3987	0.0935	0.0786	0.3804	0.0857
	250	mean	0.7319	-0.5093	0.9966	0.7292	-0.2988	0.9981
		SD	0.0535	0.2795	0.0592	0.0462	0.2706	0.0540
		$\hat{E}(I'(\theta)^2)$	0.0497	0.2564	0.0567	0.0474	0.2445	0.0519
	500	mean	0.7332	-0.4629	0.9951	0.7384	-0.2398	0.9997
		SD	0.0371	0.1673	0.0404	0.0337	0.1548	0.0373
		$\hat{E}(I'(\theta)^2)$	0.0333	0.1728	0.0386	0.0319	0.1751	0.0354

Table 2.1: Standard errors for the Lindley walk with an AR(1) $\{C_n\}$. The results report the sample standard deviation (SD) of the parameter estimators from the 200 independently generated series, and the average of the 200 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs. Both standard errors roughly agree.

Chapter 3

Discussion

This dissertation investigated Lindley random walks (storage models) in the case where the change process driving the walk is strictly stationary. This essentially extends Lindley process inference to time series settings.

First, the dissertation established the asymptotic mathematical properties of Lindley processes with correlated changes, providing a streamlined analysis. We then investigated the Markov (or lack thereof) structure of the Lindley process. The dissertation then turned to statistical estimation issues, deriving the model's likelihood function in the case of a Gaussian AR(p) change process. Because of the complexity of the resulting expression, a particle filtering method of likelihood approximation was investigated that partitioned the series into segments where the change process was either recoverable or censored. A simulation study showed that the estimation procedure works well; accurate standard errors for the parameters were even achieved.

Several directions for future research are apparent. Queueing applications involving (1.37) would need to move away from a Gaussian $\{C_n\}$. Here, one may desire the $\{I_n\}$ and $\{S_n\}$ processes to be stationary but with exponentially distributed marginal distributions. A copula way to do this for the service times $\{S_n\}$ is to take a stationary Gaussian process $\{Z_n\}$, standardized so that $E[Z_n] \equiv 0$ and $\mathrm{Var}(Z_n) \equiv 1$, and set

$$S_n = F^{-1}(\Phi(Z_n)),$$

where $F^{-1}(x) = -\ln(x)/\eta$ is the inverse of the exponential cumulative distribution function with

mean $\eta > 0$ (non-exponential distributions can also be made). Another extension involves inventory counts. Here, the process would be count valued, with $\{I_n\}$ and $\{S_n\}$ taking on count marginal distributions such as Poisson.

Applications involving environmental quantities such as daily snow or frozen soil depths may require periodic versions of the walk. Here, $\{C_n\}$ would be stationary in a periodic sense, driven by weather dynamics. This is the case in [19], where trend components are also put into the modeling procedure. One would need to investigate periodic stability of the walk and develop models that account for seasonal features and trends.

Bibliography

[1] Ishwar V Basawa, U Narayan Bhat, and Robert Lund. Maximum likelihood estimation for single server queues from waiting time data. *Queueing Systems*, 24(1):155–167, 1996.

[2] Andreas Brandt, Peter Franken, and Bernd Lisek. *Stationary Stochastic Models*. Wiley, New York, 1990.

[3] Peter J Brockwell and Richard A Davis. *Time Series: Theory and Methods*. Springer, New York, 2 edition, 1991.

[4] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 2002.

[5] Zongwu Cai. Estimating a distribution function for censored time series data. *Journal of Multivariate Analysis*, 78(2):299–318, 2001.

[6] George Casella and Roger L Berger. *Statistical Inference*. Cengage Learning, 2021.

[7] Petar M Djurić and Mónica F Bugallo. Particle filtering for high-dimensional systems. In *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 352–355. IEEE, 2013.

[8] Kerry W Fendick, Vikram R Saksena, and Ward Whitt. Dependence in packet queues. *IEEE Transactions on Communications*, 37(11):1173–1183, 1989.

[9] Kerry W Fendick, Vikram R Saksena, and Ward Whitt. Investigating dependence in packet queues with the index of dispersion for work. *IEEE Transactions on Communications*, 39(8):1231–1244, 1991.

[10] Alan Genz and Giang Trinh. Numerical computation of multivariate normal probabilities using bivariate conditioning. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 289–302. Springer, New York, 2016.

[11] Flávio B Gonçalves and Pedro Franklin. On the definition of likelihood function. *arXiv preprint arXiv:1906.10733*, 2019.

[12] Zifei Han and Victor De Oliveira. gckrig: An r package for the analysis of geostatistical count data using gaussian copulas. *Journal of Statistical Software*, 87(1):1–32, 2018.

[13] Harry Heffes. A class of data traffic processes – covariance function characterization and related queuing results. *Bell System Technical Journal*, 59(6):897–929, 1980.

[14] Harry Heffes and David Lucantoni. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, 4(6):856–868, 1986.

[15] Gang Uk Hwang and Khosrow Sohraby. On the exact analysis of a discrete-time queueing system with autoregressive inputs. *Queueing Systems*, 43(1):29–41, 2003.

[16] Faouzi Kamoun. The discrete-time queue with autoregressive inputs revisited. *Queueing Systems*, 54(3):185–192, 2006.

[17] David G Kendall. Some problems in the theory of queues. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2):151–173, 1951.

[18] Song-Hee Kim and Ward Whitt. Are call center and hospital arrivals well modeled by nonhomogeneous poisson processes? *Manufacturing & Service Operations Management*, 16(3):464–480, 2014.

[19] J. Lee, R. Lund, J. Woody, and Y. Xu. Trend assessment for daily snow depths with changepoint considerations. *Environmetrics*, 31(1):2580–2591, 2020.

[20] David V Lindley. The theory of queues with a single server. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pages 277–289. Cambridge University Press, 1952.

[21] Torgny Lindvall. *Lectures on the Coupling Method*. Wiley, New York, 2002.

[22] Attila Lovas and Miklós Rásonyi. Ergodic theorems for queuing systems with dependent interarrival times. *Operations Research Letters*, 49(5):682–687, 2021.

[23] Robert M Loynes. The stability of a queue with non-independent inter-arrival and service times. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 58, pages 497–520. Cambridge University Press, 1962.

[24] Guido Masarotto and Cristiano Varin. Gaussian copula regression in r. *Journal of Statistical Software*, 77(8):1–26, 2017.

[25] Christian Naesseth, Fredrik Lindsten, and Thomas Schon. Nested sequential Monte Carlo methods. In *International Conference on Machine Learning*, pages 1292–1301. PMLR, 2015.

[26] Jung Wook Park, Marc G Genton, and Sujit K Ghosh. Censored time series analysis with autoregressive moving average models. *Canadian Journal of Statistics*, 35(1):151–168, 2007.

[27] Patrick Rebeschini and Ramon Van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.

[28] Richard J Rossi. *Mathematical Statistics: An Introduction to Likelihood Based Inference*. John Wiley & Sons, 2018.

[29] Kotikalapudi Sriram and Ward Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, 4(6):833–846, 1986.

[30] Dag Tjøstheim and Jostein Paulsen. Bias of some commonly-used time series estimates. *Biometrika*, 70(2):389–399, 1983.

[31] Yung Liang Tong. *The Multivariate Normal Distribution*. Springer, New York, 2012.

[32] Maria Vlasiou. Lindley-type recursions. pages PhD Dissertation, Eindhoven University of Technology, 2006.

[33] Jonathan Woody, Robert Lund, Andrew J Grundstein, and Thomas L Mote. A storage model approach to the assessment of snow depth trends. *Water Resources Research*, 45(10):W10426, 2009.

[34] Scott L Zeger and Ron Brookmeyer. Regression analysis with censored autocorrelated data. *Journal of the American Statistical Association*, 81(395):722–729, 1986.