CHAPTER 2

STOCHASTIC PROCESSES AND TIME SERIES MODELLING

[*This chapter is based on the lectures of Dr. K.K. Jose, Department of Statistics, St. Thomas College, Pala, M.G. University, Kottayam.*]

2.0. Introduction

In this Chapter we discuss some elementary theory of Stochastic Processes and Time Series Modelling. Stochastic processes are introduced in Section 2.1. Some modern concepts in distribution theory which are of frequent use in this chapter are discussed Section 2.2. Section 2.3 deals with stationary time series models. In Section 2.4,we consider a structural relationship and some new autoregressive models. Section 2.5 deals with tailed processes. In section 2.6, semi-Weibull time series models with minification structure are discussed.

2.1. Stochastic Processes

The theory of stochastic processes is generally regarded as the dynamic part of Probability Theory, in which one studies a collection of random variables indexed by a parameter. One is observing a stochastic process whenever one examines a system developing in time in a manner controlled by probabilistic laws. In other words, a Stochastic Process can be regarded as an empirical abstraction of a phenomenon developing in nature according to some probabilistic rules.

If a scientist is to take account of the probabilistic nature of the phenomenon with which he is dealing, he should undoubtedly make use of the theory of stochastic processes. The scientist making measurements in his laboratory, the

meteorologist attempting to forecast weather, the control systems engineer designing a servomechanism, the electrical engineer designing a communication system, the hardware engineer developing a computer network, the economist studying price fluctuations and business cycles, the seismologist studing earthquake vibrations, the neurosurgeon studying brainwave records, the cardiologist studying the electro cardiogram etc. are encountering problems to which the theory of stochastic processes can be applied. Financial modelling and insurance mathematics are emerging areas where the theory of stochastic processes is widely used.

Examples of stochastic processes are provided by the generation sizes of populations such as a bacterial colony, life length of items under different renewals, service times in a queuing system, waiting times in front of a service counter, displacement of a particle executing Brownian motion, number of events during a particular time interval, number of deaths in a hospital on different days, voltage in an electrical system during different time instants, maximum temperature in a particular place on different days, deviation of an artificial satellite from its stipulated path at each instant of time after its launch, the quantity purchased of a particular inventory on different days etc. Suppose that a scientist is observing the trajectory of a satellite after its launch. At random time intervals, the scientist is observing whether it is deviating from the designed path or not and also the magnitude of the deviation.

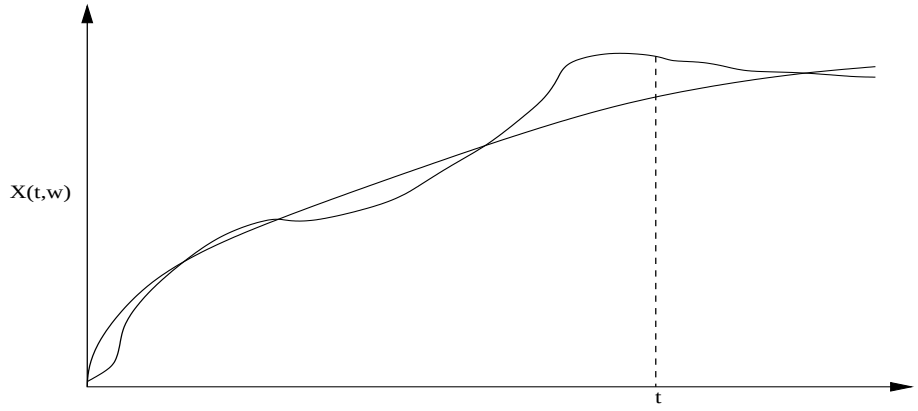

Figure 2.1

Let $X(t, w)$ denote the altitude of the satellite from sea-level at time t where w is the outcome associated with the random experiment. Here the random experiment is noting the weather conditions with regard to temperature, pressure, wind velocity, humidity etc. These outcomes may vary continuously just like that of a random experiment. Hence $\{X(t, w); t \in T; w \in \Omega\}$ gives rise to a stochastic process.

Thus a stochastic process is a family of random variables indexed by a parameter t, taking values from a set T called the index set or parameter space. It may be denoted by $\{X(t, w); t \in T, w \in \Omega\}$. A more precise definition may be given as follows.

Definition 2.1.1. A stochastic process is a family of indexed random variables $\{X(t, w); t \in T; w \in \Omega\}$ defined on a probability space (Ω, β, P) where T is an arbitrary set.

There are many ways of visualizing a stochastic process.

(i) For each choice of $t \in T, X(t, w)$ is a random variable.

(ii) For each choice of $\omega \in \Omega, X(t, w)$ is a function of t.

(iii) For each choice of w and $t, X(t, w)$ is a number.

(iv) In general it is an ensemble (family) of functions $X(t, w)$ where t and w can take different possible values.

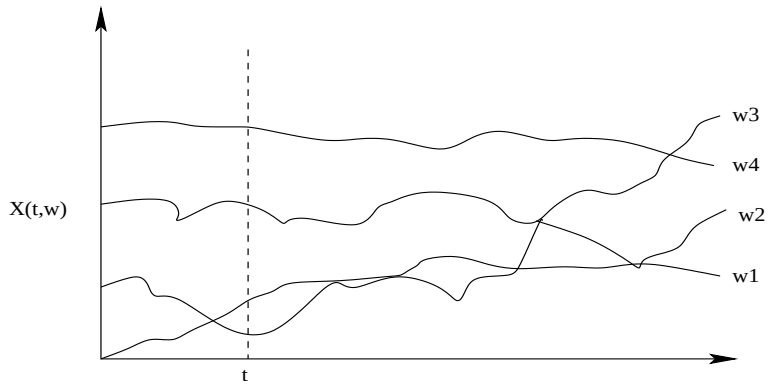

Figure 2.2

Hereafter we shall use the notation $X(t)$ to represent a stochastic process, omitting w, as in the case of random variables. It is convention to use X_n and

$X(t)$ according as the indexing parameter is discrete or continuous.

The values assumed by the r.v. (random variable) $X(t)$ are called states and the set of all possible values of $X(t)$, is called the state space of the process and is denoted by S. The state space can be discrete or continuous. When S is discrete, by a proper labeling, we can take the state-space as the set of natural numbers namely $N = \{1, 2, \cdots\}$. It may be finite or infinite.

The main elements distinguishing stochastic processes are the nature of the state space S and parameter space T, and the dependence relations among the random variables $X(t)$. Accordingly there are four types of processes.

Type 1: Discrete parameter discrete processes

In this case both S and T are discrete. Examples are provided by the number of customers reported in a bank counter on the n^{th} day, the n^{th} generation size of a population, the number of births in a hospital on the n^{th} day etc. There may be multidimensional processes also. For example consider the process (X_n, Y_n) where X_n and Y_n are the number of births and deaths in a municipality on the n^{th} day.

Type 2: Continuous parameter discrete processes

In this case T is continuous and S is discrete. Examples constitute the number of persons in a queue at time t, the number of telephone calls during $(0, t)$, the number of vehicles passing through a specific junction during $(0, t)$ etc.

Type 3: Discrete parameter continuous processes

In this case T is discrete and S is continuous. Examples are provided by the renewal time for the n^{th} renewal, life length of the n^{th} renewed bulb, service time for the n^{th} customer, waiting time on the n^{th} day to get transport, the maximum temperature in a city on the n^{th} day etc.

Type 4: Continuous parameter continuous processes

In this case both T and S are continuous. Examples are constituted by the voltage in an electrical system at time t, the blood pressure of a patient at time t, the ECG level of a patient at time t, the displacement of a particle undergoing

Brownian motion at time t, the speed of a vehicle at time t, the altitude of a satellite at time t, etc.

For more details see Karlin and Taylor (2002), Papoulis (2000), Medhi (2004), Medhi (2006). Feller (1996) gives a good account of infinite divisible distributions. Ross (2002) gives a good description of stochastic processes and their applications. Medhi (2004) gives a good introduction to the theory and application of stochastic processes.

Consider a computer system with jobs arriving at random points in time, queuing for service, and departing from the system after service completion. Let N_k be the number of jobs in the system at the time of departure of the k^{th} customer (after service completion). The stochastic process $\{N_k; k = 1, 2, \cdots\}$ is a discrete-parameter, discrete-state process. A realization of this process is shown in figure 2.3

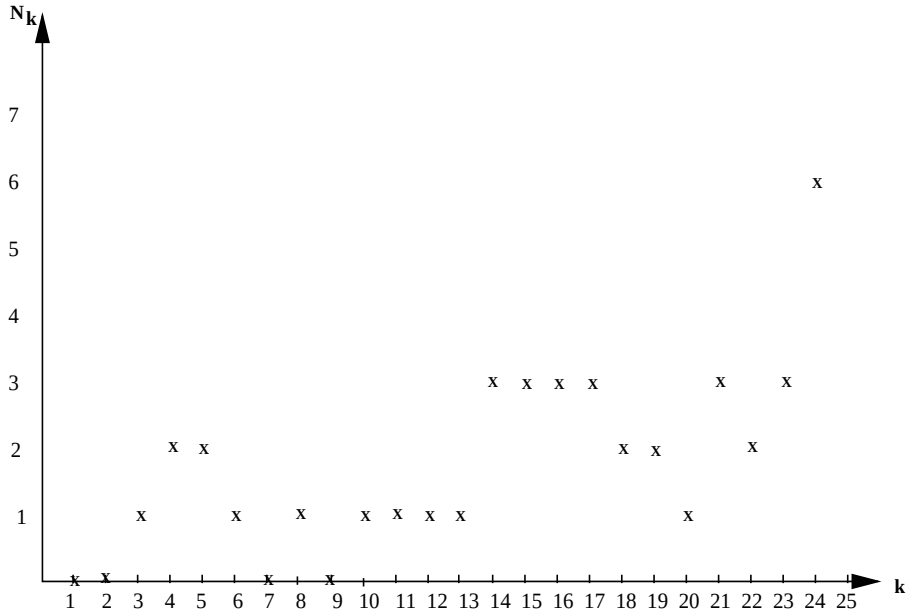

Figure 2.3

Next let $X(t)$ be the number of jobs in the system at time t. Then $\{X(t); t \in T\}$ is a continuous parameter discrete-state process. A realization is given in figure 2.4.

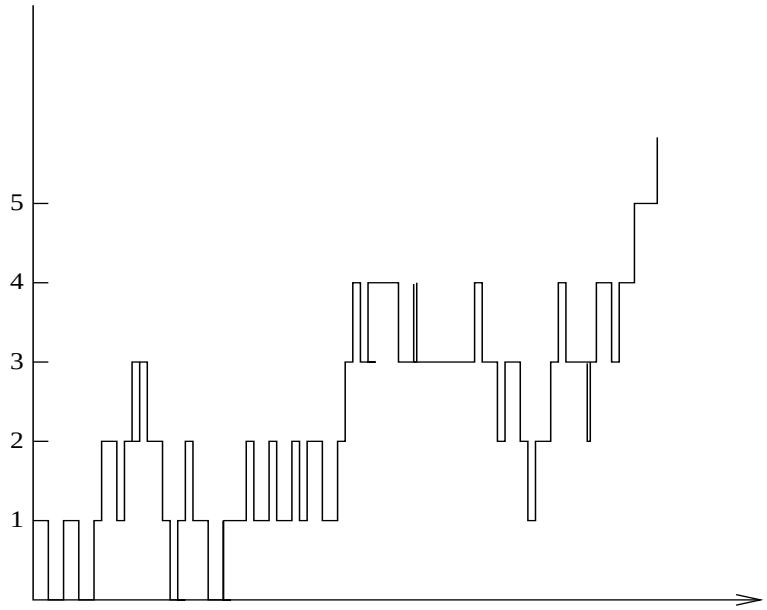

Figure 2.4

Let W_k be the time that the k^{th} customer has to wait in the system before receiving service. Then $\{W_k; k \in T\}$ is a discrete-parameter, continuous-state process, see figure 2.5

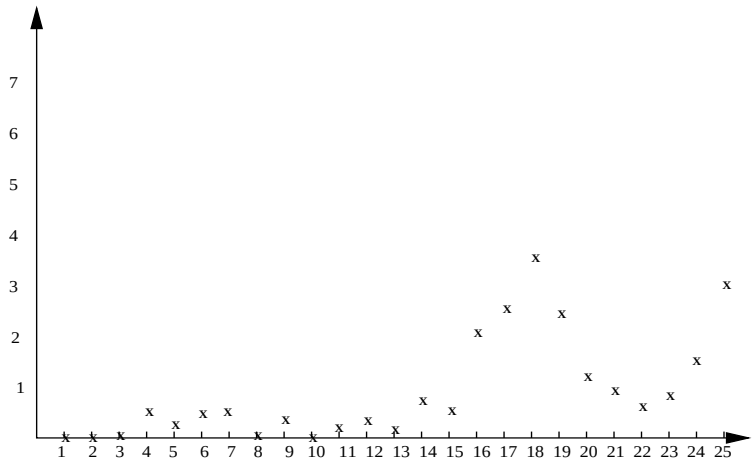

Figure 2.5

Finally, let $Y(t)$ be the cumulative service requirement of all jobs in the system at time t. Then $\{Y(t)\}$ is a continuous parameter continuous-state proces, see figure 2.6.

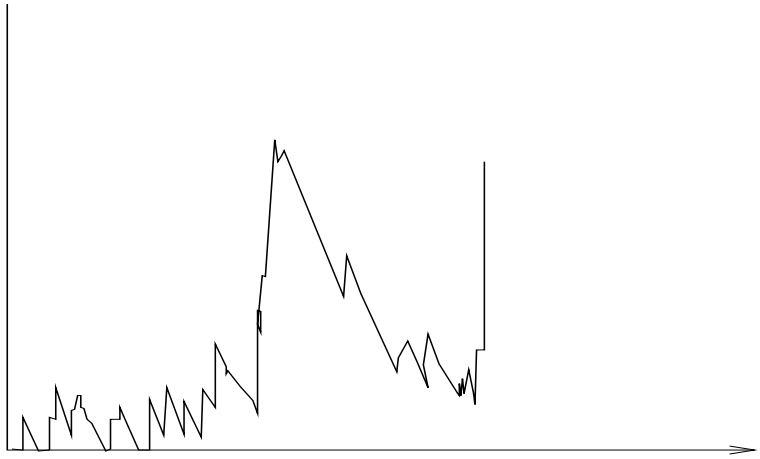

Figure 2.6

2.1.1. Classical types of stochastic processes

We now describe some of the classical types of stochastic processes characterized by different dependence relationships among $X(t)$.

2.1.2. Processes with stationary independent increments

Consider a stochastic process $\{X(t); t \in T\}$ where $T = [0, \infty)$. Then the process $\{X(t)\}$ is called a process with independent increments if the random variables $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent for all choices of $t_0, t_1, \cdots, t_n$ such that $t_0 < t_1 < \cdots < t_n$.

If the distribution of the increments $X(t_i + h) - X(t_i)$ depends only on h, the length of the interval and not on the particular time t_i, then the process is said to have stationary increments. Hence for a process with stationary increments, the distributions of the increments $X(t_0+h)-X(t_0), X(t_1+h)-X(t_1), X(t_2+h)-X(t_2), \cdots$ etc are the same and depend only on h, irrespective of the time points $t_0, t_1, \cdots$

If a process $\{X(t)\}$ has both independent and stationary increments, then it is called a process with stationary independent increments.

Result : If a process $\{X_t; t \in T\}$ has stationary independent increments and has finite mean, then $E(X_t) = m_0 + m_1 t$ where $m_0 = E(X_0)$ and $m_1 = E(X_1) - m_0$, E denoting the expected value.

2.1.3. Stationary processes

A stochastic process $\{X_t\}$ is said to be stationary in the strict sense (SSS) if the joint distribution function of the families of the random variables $[X_{t_1+h}, \cdots X_{t_n+h}]$ and $[X_{t_1}, X_{t_2}, \cdots, X_{t_n}]$ are the same for all $h > 0$ and arbitrary selections $t_1, t_2, \cdots, t_n$ from T. This condition asserts that the process is in probabilistic equilibrium and that the particular times at which one examines the process are of no relevance. In particular the distribution of X_t is the same for each t.

Thus stationarity of a process implies that the probabilistic structure of the process is invariant under translation of the time axis. Many processes encountered in practice exhibit such a characteristic. So, stationary processes are appropriate models for describing many phenomena that occur in communication theory, astronomy, biology, economics etc.

However strict sense stationarity is seldom observed in practice. Moreover, many important questions relating to a stochastic process can be adequately answered in terms of the first two moments of the process. Therefore we relax the condition of strict sense stationarity to describe weak sense stationarity (WSS), also known as wide sense stationarity.

A Stochastic process $\{X_t\}$ is said to be wide sense stationary if its first two moments (mean function and variance function) are finite and independent of t and the covariance function $\mathrm{Cov}(X_t, X_{t+s})$ is a function only of s, the time difference, for all t. Such processes are also known as covariance stationary or second order stationary processes. A process, which is not stationary, in any sense, is said to be evolutionary.

2.1.4. Gaussian processes and stationarity

If a process $\{X_t\}$ is such that the joint distribution of $(X_{t_1}, X_{t_2}, X_{t_n})$ for all $t_1, t_2, \cdots, t_n$ is multivariate normal, then $\{X_t\}$ is called a Gaussian (normal) process. For a Gaussian process weak sense stationarity and strict sense stationarity are identical. This follows from the fact that a multivariate normal distribution is

completely determined by its mean vector and variance-covariance matrix. Here we need only mean, variance and covariance functions. In other words, if a Gaussian process $\{X_t\}$ is covariance stationary, then it is strictly stationary and vice versa.

Example 2.1.1. Let $\{X_n; n \geq 1\}$ be uncorrelated random variables with mean 0 and variance 1. Then

$$\text{Cov}(X_n, X_m) = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} .$$

Hence $\text{Cov}(X_n, X_m)$ is a function of $n - m$ and so the process is covariance stationary. If X_n are identically distributed also, then the process is strictly stationary.

Example 2.1.2. Consider the Poisson process $\{X(t)\}$ where

$$P[X(t) = n] = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}; n = 0, 1, \cdots .$$

Clearly,

$$E[X(t)] = \lambda t$$
$$\text{Var}[X(t)] = \lambda t \text{ which depends on } t$$

Therefore the process is not stationary. It is evolutionary.

Example 2.1.3. Consider the process $\{X(t)\}$ where $X(t) = A_1 + A_2 t$ where A_1, A_2 are independent r.v.'s with $E(A_i) = a_i, \text{Var}(\text{A}_\text{i}) = \sigma_\text{i}^2, \text{i} = 1, 2$. Obviously

$$E[X(t)] = a_1 + a_2 t$$
$$\text{Var[X(t)]} = \sigma_1^2 + \sigma_2^2 t^2$$
$$\text{Cov[X(s), [X(t)]} = \sigma_1^2 + st\sigma_2^2.$$

These are functions of t and hence the process is evolutionary.

Example 2.1.4. Consider the process $\{X(t)\}$ where $X(t) = A\cos wt + B\sin wt$, where A and B are uncorrelated r.v.'s with mean 0 and variance 1 and w is a positive constant.

In this case $E[X(t)] = 0$ and $\text{Var[X(t)]} = 1$, $\text{Cov[X(t), X(t + h)]} = \cos(\text{hw})$. Hence the above process is covariance stationary. This process is called a sinusoidal process.

Example 2.1.5. Consider the process $\{X(t)\}$ such that

$$P[X(t)=n]=\begin{cases}\frac{(at)^{n-1}}{(1+at)^{n+1}}; n=1,2,\cdots\\ \frac{(at)}{(1+at)}; n=0.\end{cases}$$

Obviously

$$\begin{aligned}E[X(t)] &= \sum_{n=0}^{\infty} nP[X(t)=n]\\ &= \frac{1}{(1+at)^2}\sum_{n=1}^{\infty} n\frac{(at)^{n-1}}{(1+at)} = 1\\ E[X^2(t)] &= \sum_{n=1}^{\infty} n^2\frac{(at)^{n-1}}{(1+at)^{n-1}}\\ &= \frac{at}{(1+at)^3}\left\{\sum_{n=2}^{\infty} n(n-1)\frac{(at)^{n-2}}{(1+at)}\right\}\\ &\quad + \sum_{n=1}^{\infty} n\frac{(at)^{n-1}}{(1+at)^{n+1}}\\ &= 2at+1,\end{aligned}$$

which is a function of t. Hence the process is not stationary. It is an evolutionary process.

Example 2.1.6. Consider the Bernoulli process described below.

Consider a sequence of independent Bernoulli trials with outcomes as success and failure. Let

$$X_n=\begin{cases}1 & \text{if the outcome is a success}\\ 0 & \text{otherwise.}\end{cases}$$

Then the process $\{X_n; n\geq 1\}$ has states 0 and 1 and the process is called a Bernoulli process. Let us define $\{Y_n\}$ by $Y_n=0$ for $n=0$ and $Y_n=X_1+\cdots+X_n, n\geq 1$. Then the process $\{Y_n; n\geq 0\}$ has the set of non-negative integers as the state space. The Y_{n_i} is binomially distributed with $P[Y_n=k]={}_nC_k p^k(1-p)^{n-k}; k=0,1,2,\cdots n$; where p is the probability of success in a trial.

Example 2.1.7. (The random telegraph signal process)

Let $\{N(t), t \geq 0\}$ denote a Poisson process, and let X_0 be independent of this process, and be such that $P(X_0 = 1) = P(X_0 = -1) = \frac{1}{2}$. Define $X(t) = X_0(-1)^{N(t)}$. Then $\{X(t); t \geq 0\}$ is called a random telegraph signal process. In this case $P[N(t) = k] = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^k}{k!}$ for $k = 1, 2, \cdots$. Clearly

$$\begin{aligned}
E[X(t)] &= E[X_0(-1)^{N(t)}] \\
&= E[X_0]E[(-1)^{N(t)}] = 0 \\
\mathrm{Cov}[X(t), X(t+s)] &= E[X(t)X(t+s)] \\
&= E[X_0^2(-1)^{N(t)+N(t)+s}] \\
&= E[X_0^2].E[(-1)^{2N(t)+N(t)+s+N(t)-N(t)}] \\
&= 1E[(-1)^{2N(t)}(-1)^{N(t+s)-N(t)}] \\
&= E[(-1)^{2N(t)}]E[(-1)^{N(t+s)}] \\
&= 1E[(-1)^{N(s)}] \\
&= \sum_{k=0}^{\infty}(-1)^k\frac{\mathrm{e}^{\lambda s}(\lambda s)^k}{k!} \\
&= \mathrm{e}^{-2\lambda s}; s \geq 0.
\end{aligned}$$

Also

$$\mathrm{Var}[X(t)] = 1 < \infty.$$

Hence the above process is covariance stationary.

For an application of the above random telegraph signal, consider a particle moving at a constant unit velocity along a straight line and suppose that collisions involving this particle occur at Poisson rate λ. Also suppose that each time the particle suffers from a collision, it reverses direction. If X_0 represents the initial velocity of the particle, then its velocity at time t is given by $X(t) = X_0(-1)^{N(t)}$. If we take $D(t) = \int_0^t x(\mathrm{s})\mathrm{ds}$, then $D(t)$ represents the displacement of the particle during $(0, t)$. It can be shown that $\{D(t); t \geq 0\}$ is also a weakly stationary process.

Example 2.1.8. Consider an Autoregressive Process $\{X_n\}$ where $X_0 = Z_0$ and $X_n = \rho X_{n-1} + Z_n; n \geq 1, |\rho| < 1$ where $Z_0, Z_1, Z_2, \cdots$ are uncorrelated random variables with $E(Z_n) = 0; n \geq 0$ and

$$\text{Var}(Z_n) = \begin{cases} \sigma^2; & n \geq 1 \\ \frac{\sigma^2}{1-\rho^2}; & n = 0. \end{cases}$$

Then

$$\begin{aligned} X_n &= \rho X_{n-1} + Z_n \\ &= \rho(\rho X_{n-2} + Z_{n-1} + Z_n \\ &= \rho^2 X_{n-2} + \rho Z_{n-1} + Z_n \\ &= \sum_{i=0}^{n} \rho^{n-i} Z_i. \end{aligned}$$

Therefore

$$\begin{aligned} E(X_n) &= 0 \\ \text{Cov}[X_n, X_{n+m}] &= \text{Cov}[\sum_{i=0}^{n} \rho^{n-1} Z_i, \sum_{i=0}^{n+m} \rho^{n+m-i} Z_i \\ &= \sum_{i=0}^{n} \rho^{n-1} \rho^{n+m-i} \text{Cov}(Z_i, Z_i) \\ &= \sigma^2 \rho^{2n+m} \left[\frac{1}{1+\rho^2} + \sum_{i=1}^{n} \rho^{-2i} \right] \\ &= \frac{\sigma^2 \rho^m}{1-\rho^2}. \end{aligned}$$

Therefore this process is also covariance stationary.

Now we consider a special type of Gaussian Process, which is stationary in both senses and has a wide range of applications.

2.1.5. Brownian processes

We consider a symmetric random walk in which in each time unit there is chance for one unit step forward or backward. Now suppose that we speed up this process by taking smaller and smaller steps in smaller and smaller time intervals. In the limit we obtain the Brownian motion process. It is also known as the Wiener process, after Wiener who developed this concept in a series of papers

from 1918 onwards. Actually it originated in Physics, as the notion associated with the random movements of a small particle immersed in a liquid or gas. This was first discovered by the British botanist Robert Brown. The process can be more precisely developed as follows:

Suppose that, in the random walk, in each time interval of duration Δt we take a step of size Δx either to the left or to the right with equal probabilities. If we let $X(t)$ denote the position at time t, then

$$X(t) = \Delta x\,[X_1 + \cdots + X_{(\frac{t}{\Delta t})}]$$

where

$$X_n = \begin{cases} +1 & \text{if the } i^{th} \text{ step is to the right} \\ -1 & \text{if it is to the left} \end{cases}$$

and $\frac{t}{\Delta t}$ is the integer part of $\frac{t}{\Delta t}$. We assume that $X_i's$ are independent with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Since $E(X_i) = 0, \text{Var}(X_i) = 1$ we have $E[X(t)] = 0, \text{Var}[X(t)] = (\Delta x)^2[\frac{t}{\Delta t}]$

Now we consider the case when $\Delta x \to 0$ and $\Delta t \to 0$ in such a way that $E[X(t)] = 0$ and $\text{Var}[X(t)] \to \sigma^2 t$. The resulting process $\{X(t)\}$ is such that $X(t)$ is normally distnbuted with mean σ and variance $\sigma^2 t$, and has independent, stationary increments. This leads us to the formal definition of a Brownian motion process.

Definition 2.1.2. A stochastic process $\{X(t); t \geq 0\}$ is said to be a Brownian motion process if (i) $X(0) = 0$ (ii) $\{X(t)\}$ has stationary independent increments (iii) for every $t > 0, X(t)$ is normally distributed with mean 0 and variance $\sigma^2 t$.

When $\sigma = 1$, the process is called a standard Brownian motion. Any Brownian motion $X(t)$ can be converted to a standard Brownian motion by taking $B(t) = \frac{X(t)}{\sigma}$. If $\{B(t)\}$ is a standard Brownian motion and $X(t) = \sigma B(t) + \mu t$, then $X(t)$ is normally distributed with mean μt and variance $t\sigma^2$. Then $\{X(t); t \geq 0\}$ is called a Brownian motion with drift coefficient μ.

If $\{X(t); t \geq 0\}$ is a Brownian motion process with drift coefficient μ and variance parameter $\sigma^2 t$, then $\{Y(t); t \geq 0\}$ where $Y(t) = \exp[X(t)]$ is a called a geometric Brownian motion process. It is useful in modeling of stock prices over time when the percentage changes are independent and identically distributed.

If $\{X(t); t \geq 0\}$ is a Brownian motion process then each of $X(t_l), X(t_2), \cdots$ can be expressed as a linear combination of the independent normal random variables $X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \cdots X(t_n) - X(t_{n-1})$. Hence it follows that a Brownian motion is a Gaussian process.

Since a multivariate normal distribution is completely determined by the marginal mean values and covariance values, it follows that a standard Brownian motion could also be defined as a Gaussian process having $E[X(t)] = 0$ and for $s \geq t$,

$$\begin{aligned}
\text{Cov}[X(s), X(t)] &= \text{Cov}[X(s), X(s) + X(t) - X(s)] \\
&= \text{Cov}[X(s), X(s)] + \text{Cov}[X(s), X(t) - X(s)] \\
&= \text{Var}[X(s)] \\
&= s\sigma^2.
\end{aligned}$$

Let $\{X(t); t \geq 0\}$ be a standard Brownian motion process and consider the process values between 0 and 1 conditional on $X(1) = 0$. Consider the conditional stochastic process, $\{X(t); 0 \geq t \geq 1 | X(1) = 0\}$. Since this conditional distribution is also multivariate normal it follows that this conditional process is a Gaussian process. This conditional process is known as the Brownian bridge.

Brownian motion theory is a major topic in fluid dynamics and has applications in aeronautical engineering in the designing of aeroplanes, submarines, satellites, space crafts etc. It also finds applications in financial modelling.

2.1.6. Markov chains

An elementary form of dependence between values of X_n in successive transitions, was introduced by the celebrated Russian probabilist A.A. Markov, and is known as Markov dependence. Markov dependence is a form of dependence

which states that X_{n+1} depends only on X_n when it is known and is independent of $X_{n-1}, X_{n-2}, \cdots, X_0$. This implies that the future of the process depends only on the present, irrespective of the past. This property is known as Markov property. In probabilistic terms, the Markov property can be stated as

$$\begin{aligned} P[X_{n+1}, = i_{n+1} | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i_n] \\ = P[X_{n+1} = i_{n+1} | X_n = i_n] \end{aligned}$$

for all states $i_0, i_1, \cdots, i_{n+1}$ and for all n. This is called Markov dependence of the first order.

A Stochastic process $\{X_n\}$ with discrete state space and discrete parameter space is called a markov chain if for all states $i, j, i_0, i_1, \cdots, i_{n-1}$ we have

$$\begin{aligned} P[X_{n+1} = j | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i] \\ = P\{X_{n+1} = j | X_n = i] \qquad \text{for all } n. \end{aligned}$$

The probability that the system is in state j at the end of $(n+1)$ transitions given that the system was in state i at the end of n transitions is denoted by $p_{ij}^{(1)}$ and is called a one-step transition probability. In general this probability depends on i, j and n. If these probabilities are independent of n, we say that the Markov chain is homogeneous and has stationary transition probabilities. Here we consider only such chains. Thus

$$p_{ij}^{(1)} = P[X_{n+1} = j | X_n = i]$$

In a similar manner we can consider m-step transition pribabilities denoted by $p_{ij}^{(m)}$ where

$$p_{ij}^{(m)} = P[X_{n+m} = j | X_n = i]$$

If the state space of a Markov chain consists of only a finite number of states, it is called a finite Markov chain. Otherwise we call it an infinite Markov chain.

The square matirx P consting of the elements $p_{ij}^{(1)}$ for all possible states i and j is called one -step transition probability matrix of the chain. Therefore

$$P = [p_{ij}^{(1)}].$$

Similarly the square matrix $P^{(m)}$ consisting of the elements $p_{ij}^{(m)}$ for all possible values of the states i and j is called the m-step transition matrix of the chain. Hence

$$P^{(m)} = [p_{ij}^{(m)}].$$

Obviously we have $P^{(1)} = P$ and

$$p_{ij}^{(m)} \geq 0 \qquad \text{and} \qquad \sum_j p_{ij}^{(m)} = 1.$$

Now we consider $p_j^{(0)} = P[X_0 = j]$. It may be noted that $p_j^{(0)}$ describes the probability distribution of X_0. The vector $p^{(0)} = (p_0^{(0)}, p_1^{(0)}, \cdots p_j^{(0)} \cdots)$ is called the initial probability vector.

Similarly $p_j^{(n)} = P[X_n = j]$ gives the probability distribution of X_n. The vector $p^{(n)} = (p_0^{(n)}, p_1^{(n)}, \cdots p_j^{(n)} \cdots)$ is called the n-step absolute probability vector.

Theorem 2.1.1. *A Markov chain is completely defined by its one-step transition probability matrix and the initial probability vector.*

Proof 2.1.1. Consider

$$\begin{aligned}
&P[X_0 = i|X_1 = j, X_2 = k, \cdots, X_{n-1} = r, X_n = s] \\
&\quad = P[X_0 = i]\, P[X_1 = j|X_0 = i]\, P[X_2 = k|X_0 = i, X_1 = j] \\
&\qquad \cdots P\{X_n = s|X_0 = i, \cdots, X_{n-1} = r\} \\
&\quad = P(X_0 = i)P(X_1 = j|X_0 = i)\, P(X_2 = k|X_1 = j) \cdots P(X_n = s|X_{n-1} = r) \\
&\quad = p_i^{(0)} p_{ij}^{(1)} \cdots p_{rs}^{(1)}
\end{aligned}$$

This shows that any finite dimensional joint distribution for the chain can be obtained in terms of the initial probabilities and one-step transition probabilities, and this establishes the thorem.

Theorem 2.1.2. *(Chapman-Kolmogorov Equations)*

The transition probabilities of Markov chains satisfy the equation

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)} \qquad \textit{where} \quad p_{ij}^{(0)} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

or equivalently,

$$P^{(n)} = P^n \quad \textit{and} \quad P^{(m+n)} = P^{(m)} P^{(n)}$$

(i) **Computation of absolute probabilities**

Consider

$$\begin{aligned} p_j^{(n)} &= P[X_n = j] \\ P(X_n = j) &= \sum_i P(X_n = j, X_0 = i) \\ &= \sum_i P(X_0 = i) P(X_n = j | X_0 = i) \end{aligned}$$

Therefore

$$p_j^{(n)} = \sum_i p_i^{(0)} p_{ij}^{(n)}$$

(ii) **Inverse transition probabilities**

The n-step inverse transition probabilities denoted by $q_{ij}^{(n)}$ is defined as

$$q_{ij}^{(n)} = P(X_m = j | X_{n+m} = i)$$

for $m \geq 0, n \geq 0$. They describe the past behaviour of the process when the present is given. But transition probabilities describe the future behaviour of the process when the present is given.

Now

$$\begin{aligned} &P(X_m = j | X_{n+m} = i) P(X_{n+m} = i) \\ &\quad = P(X_{n+m} = i | X_n = j) P(X_n = j) \end{aligned}$$

Therefore

$$P(X_m = j|X_{n+m} = i) = \frac{P(X_{n+m} = i|X_m = j)P(X_m = j)}{P(X_{n+m} = i)}$$

Hence

$$q_j^{(n)} = \frac{p_{ij}^{(n)} p_j^{(m)}}{P_i^{(n+m)}} \qquad m \geq 0$$

whenever the denominator is non zero.

(iii) **Taboo probabilities**

In this case the movement of the system to some specified states is prohibited. For example consider

$$P(X_2 = j, X_1 \neq k|X_0 = i) = P$$

(the system reaches state j at the end of 2 transitions without visiting state k given that the system started from state i).

This is usually denoted by ${}_kP_{ij}^{(2)}$. Using the Chapman-Kolmogorov equations, we have

$${}_kp_{ij}^{(2)} = \sum_{1 \neq k} p_{il}^{(1)} p_{lj}^{(1)}$$

It may be noted that ${}_kP_{ij}^{(2)}$ is different from $P[X_2 = j|X_0 = i, X_1 \neq k]$ which is equal to $P[X_2 = j|X_1 \neq k]$.

Similarly $P[X_2 = j, X_1 \neq k, l|X_0 = i]$ may be denoted by ${}_{l,k}P_{ij}^{(2)}$.
Obviously,

$${}_{l,k}P_{ij}^{(2)} = \sum_{v \neq k,l} p_{iv}^{(1)} p_{vj}^{(1)}$$

Problems relating to taboo probabilities can be solved as shown above.

Exercises 2.1.

2.1.1. Give two examples each of the four types of stochastic processes.

2.1.2. Define a stochastic process with stationary independent increments.

2.1.3. For a process with stationary independent increments show that $E(X_t) = m_0 + m_1 t$ where $m_0 = E(X_0)$ and $m_1 = E(X_1) - m_0$.

2.1.4. What is a Poisson process ? Show that it is evolutionary.

2.1.5. Give an example of a strictly stationary process.

2.1.6. Give an example of a covariance stationary process.

2.1.7. Let $\{X_n\}$ be uncorrelated r.v.'s with $E(X_n) = 0, V(X_n) = 1$. Show that $\{X_n\}$ is strictly stationary.

2.1.8. Consider a Poisson process $\{x(t)\}$ where $p[x(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$; $n = 0, 1, \cdots$ Find $E(x(t))$ and $\text{Var}(x(t))$. Is the process stationary ?

2.1.9. Consider a Poisson process $\{x(t)\}$ as above. Let x_0 be independent of $x(t)$ such that $p(x_0 = 1) = p(x_0 = -1) = \frac{1}{2}$. define $N(t) = x_0(-1)^{N(t)}$. Find $E(N(t))$ and $\text{Cov}(N(t), N(t+s))$.

2.1.10. Define a Brownian process and show that it is an approximation of the random walk process.

2.1.11. Obtain an expression for the covariance function of a Brownian motion process.

2.1.12. What is geometric Brownian motion process? Discuss its uses ?

2.1.13. Consider the numbers $1, 2, 3, 4, 5$. We select one number out of these at random and note it as X_1. Then select a number at random from $1, 2 \cdots X_1$ and denote it as X_2. The process is continued. Write down the one step and two step transition matrices of the chain $\{X_n\}$.

2.1.14. 4 white and 4 red balls are randomly distributed in two urns so that each urn contains 4 balls. At each step one ball is selected at random from each urn and the two balls are interchanged. Let X_n denote the number of white balls in the first urn at the end of the n^{th} interchange. Then write down the one-step transition matrix and the initial distribution. Also find

(i) $P[X_3 = 4|X_1 = 4]$ (ii) $P[X_2 = 3]$

(iii) $P[X_1 = 4,\ X_2 = 3,\ X_3 = 2,\ X_4 = 1]$ (iv) $P[X_1 = 3|X_3 = 4]$

2.1.15. If X_n denotes the maximum face value observed in n tosses of a balanced die with faces marked $1, 2, 3, 4, 5, 6$ write down the state space and parameter space of the process $\{X_n\}$. Also obtain the transition matrix.

2.2. Modern Concepts in Distribution Theory

2.2.1. Introduction

In this Section we discuss some modern concepts in distribution theory which will be of frequent use in this chapter.

Definition 2.2.1. Infinite divisibility.

A random variable x is said to be infinitely divisible if for every $n \in \mathbb{N}$, there exists independently and identically distributed random variables $y_{1n}, y_{2n}, \ldots, y_{nn}$ such that $x \stackrel{d}{=} y_{1n} + y_{2n} + \cdots + y_{nn}$, where $\stackrel{d}{=}$ denotes equality in distributions. In terms of distribution functions, a distribution function F is said to be infinitely divisible if for every positive integer n, there exists a distribution function F_n such that $F = \underbrace{F_n \star F_n \star \cdots \star F_n}_{n \text{ times}}$, where $\star$ denotes convolution.

This is equivalent to the existence of a characteristic function $\varphi_n(t)$ for every $n \in \mathbb{N}$ such that $\varphi(t) = [\varphi_n(t)]^n$ where $\varphi(t)$ is the characteristic function of x.

Infinitely divisible distributions occur in various contexts in the modelling of many real phenomena. For instance when modelling the amount of rain x that falls in a period of length T, one can divide x into more general independent parts from the same family. That is,

$$x \stackrel{d}{=} x_{t_1} + x_{t_2 - t_1} + \cdots + x_{T - t_{n-1}}.$$

Similarly, the amount of money x paid by an insurance company during a year must be expressible as the sum of the corresponding amounts $x_1, x_2, \ldots, x_{52}$ in

each week, That is,

$$x \stackrel{d}{=} x_1 + x_2 + \cdots + x_{52}.$$

A large number of distributions such as normal, exponential, Weibull, gamma, Cauchy, Laplace, logistic, lognormal, Pareto, geometric, Poisson, etc., are infinitely divisible. Various properties and applications of infinitely divisible distributions can be found in Laha and Rohatgi (1979) and Steutel (1979).

2.2.2. Geometric infinite divisibility

The concept of geometric infinite divisibility (g.i.d.) was introduced by Klebanov *et al.* (1984). A random variable y is said to be g.i.d. if for every $p \in (0, 1)$, there exists a sequence of independently and identically distributed random variables $x_1^{(p)}, x_2^{(p)}, \ldots$ such that

$$y \stackrel{d}{=} \sum_{j=1}^{N(p)} x_j^{(p)} \tag{2.2.1}$$

and

$$P\{N(p) = k\} = p(1-p)^{k-1}, \qquad k = 1, 2, \cdots$$

where $y, N(p)$ and $x_j^{(p)}$, $j = 1, 2, \ldots$ are independent. The relation (2.2.1) is equivalent to

$$\begin{aligned} \varphi(t) &= \sum_{j=1}^{\infty} [g(t)]^j \; p(1-p)^{j-1} \\ &= \frac{pg(t)}{1-(1-p)g(t)} \end{aligned}$$

where $\varphi(t)$ and $g(t)$ are the characteristic functions of y and $x_j^{(p)}$ respectively.

The class of g.i.d. distributions is a proper subclass of infinitely divisible distributions. The g.i.d. distributions play the same role in 'geometric summation' as infinitely divisible distributions play in the usual summation of independent random variables. Klebanov *et al.* (1984) established that a distribution function F with characteristic function $\varphi(t)$ is g.i.d. if and only if $\exp\left\{1 - \frac{1}{\varphi(t)}\right\}$ is

infinitely divisible. Exponential and Laplace distributions are obvious examples of g.i.d. distributions. Pillai (1990b), Mohan *et al.* (1993) discuss properties of g.i.d. distributions. It may be noted that normal distribution is not geometrically infinitely divisible.

2.2.3. Bernstein functions

A C^∞–function f from $(0,\infty)$ to R is said to be completely monotone if $(-1)^n \frac{d^n f}{dx^n} \geq 0$ for all integers $n \geq 0$.

A C^∞–function f from $(0,\infty)$ to R is said to be a Bernstein function, if $f(x) \geq 0$, $x > 0$ and $(-1)^n \frac{d^n f}{dx^n} \leq 0$ for all integers $n \geq 1$. Then f is also referred to as a function with complete monotone derivative (c.m.d).

A completely monotone function is positive, decreasing and convex while a Bernstein function is positive, increasing and concave (see Berg and Forst (1975)).

Fujita (1993) established that a cumulative distribution function G with $G(0) = 0$ is geometrically infinitely divisible, if and only if G can be expressed as

$$G(x) = \sum_{n=1}^{\infty} (-1)^{n+1} W^{n*}([0,x]), \qquad x > 0$$

where $W^{n*}(dx)$ is the n-fold convolution measure of a unique positive measure $W(dx)$ such that

$$\frac{1}{f(x)} = \int_0^\infty e^{-sx} W(ds), \qquad x > 0$$

where $f(x)$ is a Bernstein function, satisfying the conditions

$$\lim_{x \downarrow 0} f(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} f(x) = \infty.$$

A distribution is said to have complete monotone derivative if its distribution function $F(x)$ is Bernstein. Pillai and Sandhya (1990) proved that the class of distributions having complete monotone derivative is a proper subclass of g.i.d. distributions. This implies that all distributions with complete monotone densities are geometrically infinitely divisible. It is easier to verify the complete monotone criterion and using this approach we can establish the geometric infinite divisibility of many distributions such as Pareto, gamma and Weibull.

The class of non-degenerate generalized gamma convolutions with densities of the form given by

$$f(x) = c\ x^{\beta-1} \prod_{j=1}^{M} (1 + c_j x)^{-r_j}, \qquad x > 0$$

is geometrically infinitely divisible for $0 < \beta \leq 1$. Similarly distributions having densities of the form

$$f(x) = cx^{\beta-1} \exp(-cx^{\alpha}); \qquad 0 < \alpha \leq 1$$

is g.i.d. for $0 < \beta \leq 1$. Also the Bondesson family of distributions with densities of the form

$$f(x) = cx^{\beta-1} \prod_{j=1}^{M} \left[1 + \sum_{k=1}^{N_j} c_{jk} x^{\alpha_{jk}} \right]^{-r_j}$$

is g.i.d. for $0 \leq \beta \leq 1$, $\alpha_{jk} \leq 1$ provided all parameters are strictly positive (see Bondesson(1992)).

2.2.4. Self-decomposability

Let $\{x_n;\ \ n \geq 1\}$ be a sequence of independent random variables, and let $\{b_n\}$ be a sequence of positive real numbers such that

$$\lim_{n\to\infty} \max_{1\leq k\leq n} P\{|x_k| \geq b_n \epsilon\} = 0 \ \text{ for every } \epsilon > 0\ .$$

Let $s_n = \sum_{k=1}^{n} x_k$ for $n \geq 1$. Then the class of distributions which are the weak limits of the distributions of the sums $b_n^{-1} s_n - a_n$; $n \geq 1$ where a_n and $b_n > 0$ are suitably chosen constants, is said to constitute class L. Such distributions are called self-decomposable.

A distribution F with characteristic function $\varphi(t)$ is called self - decomposable, if and only if, for every $\alpha \in (0, 1)$, there exists a characteristic function $\varphi_\alpha(t)$ such that $\varphi(t) = \varphi(\alpha t)\varphi_\alpha(t) \qquad$ for $\ t \in R$.

Clearly, apart from $x \equiv 0$, no lattice random variable can be self–decomposable. All non-degenerate self–decomposable distributions are absolutely continuous.

A discrete analogue of self–decomposability was introduced by Steutel and Van Harn (1979). A distribution on $N_0 \equiv \{0, 1, 2, \ldots\}$ with probability generating

function (p.g.f.) $P(z)$ is called discrete self–decomposable if and only if $P(z) = P(1-\alpha+\alpha z)P_\alpha(z)$; $|z| \le 1$, $\alpha \in (0,1)$ where $P_\alpha(z)$ is a p.g.f.

If we define $G(z) = P(1-z)$, then $G(z)$ is called the alternate probability generating function (a.p.g.f.). Then it follows that a distribution is discrete self–decomposable if and only if $G(z) = G(\alpha z)G_\alpha(z)$; $|z| \le 1$, $\alpha \in (0,1)$ where $G_\alpha(z)$ is some a.p.g.f.

2.2.5. Stable distributions

A distribution function F with characteristic function $\varphi(t)$ is stable if for every pair of positive real numbers b_1 and b_2, there exist finite constants a and $b > 0$ such that $\varphi(b_1 t)\varphi(b_2 t) = \varphi(bt)e^{iat}$ where $i = \sqrt{-1}$.

Clearly, stable distributions are in class L with the additional condition that the random variables x_n; $n \ge 1$ in Subsection 4.3.4 are identically distributed also. F is stable if and only if its characteristic function can be expressed as

$$\ln \varphi(t) = i\alpha t - c|t|^\beta[1 + i\gamma\omega(t,\beta)sgn\ t]$$

where α, β, γ are constants with $c \ge 0$, $0 < \beta \le 2$, $|\gamma| \le 1$ and

$$\omega(t,\beta) = \begin{cases} \tan\frac{\pi\beta}{2}; & \beta \ne 1 \\ \frac{2}{\pi}\ln|t|; & \beta = 1. \end{cases}$$

The value $c = 0$ corresponds to the degenerate distribution, and $\beta = 2$ to the normal distribution. The case $\gamma = 0, \beta = 1$ corresponds to the Cauchy law (see Laha and Rohatgi (1979)).

2.2.6. Geometrically strictly stable distributions

A random variable y is said to be geometrically strictly stable (g.s.s.) if for any $p \in (0,1)$ there exists a constant $c = c(p) > 0$ and a sequence of independent and identically distributed random variables $y_1, y_2, \ldots$ such that

$$y \overset{d}{=} c(p)\sum_{j=1}^{N(p)} y_j$$

where $P\{N(p) = k\} = p(1-p)^{k-1}$; $k = 1, 2, \ldots$ and $y, N(p)$ and y_j; $j = 1, 2, \ldots$ are independent.

If $\varphi(t)$ is the characteristic function of y, then it implies that

$$\varphi(t) = \frac{p\varphi(ct)}{1-(1-p)\varphi(ct)}; \qquad p \in (0,1).$$

Among the geometrically strictly stable distributions, the Laplace distribution and exponential distribution possess all moments. A geometrically strictly stable random variable is clearly geometrically infinitely divisible.

A non–degenerate random variable y is geometrically strictly stable if and only if its characteristic function is of the form

$$\varphi(t) = 1\Big/\left[1+\lambda|t|^{\alpha}\ \exp\left(-i\frac{\pi}{2}\theta\alpha\ sgn\, t\right)\right]$$

where $0 < \alpha \leq 2$, $\lambda > 0$, $|\theta| \leq \min(1, 2/\alpha - 1)$. When $\alpha = 2$, it corresponds to the Laplace distribution. Thus it is apparent that when ordinary summation of random variables is replaced by geometric summation, the Laplace distribution plays the role of the normal distribution, and exponential distribution replaces the degenerate distribution (see Klebanov *et al.*(1984)).

2.2.7. Mittag-Leffler distribution

The Mittag-Leffler distribution was introduced by Pillai (1990a) and has cumulative distribution function given by

$$F_{\alpha}(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}x^{k\alpha}}{\Gamma(1+k\alpha)}; \qquad 0 < \alpha \leq 1;\ x > 0.$$

Its Laplace transform is given by $\phi(t) = \dfrac{1}{1+t^{\alpha}}$; $0 < \alpha \leq 1$; $t \geq 0$; and the distribution may be denoted by ML(α). Here α is called the exponent. It can be regarded as a generalization of the exponential distribution in the sense that $\alpha = 1$ corresponds to the exponential distribution. The Mittag-Leffler distribution is geometrically infinitely divisible and belongs to class L. It is normally attracted to the stable law with exponent α.

If u is exponential with unit mean and y is positive stable with exponent α, then $x = u^{1/\alpha}y$ is distributed as Mittag-Leffler (α). If u is Mittag-Leffler (α) and v is exponential and u and v are independent, then $x = \dfrac{u}{v}$ is distributed as Pareto type III with survival function $\bar{F}_x(x) = P(x > x) = \dfrac{1}{1+x^{\alpha}}$; $0 < \alpha \leq 1$.

For the Mittag-Leffler distribution, $E(x^\delta)$ exists for $0 \le \delta < \alpha$ and is given by

$$E(x^\delta) = \frac{\Gamma(1-\delta/\alpha)\Gamma(1+\delta/\alpha)}{\Gamma(1-\delta)}.$$

A two parameter Mittag-Leffler distribution can also be defined with the corresponding Laplace transform $\phi(t) = \frac{\lambda^\alpha}{\lambda^\alpha + t^\alpha}$; $0 < \alpha \le 1$. It may be denoted by ML(α, λ).

Jayakumar and Pillai (1993) considered a more general class called semi–Mittag-Leffler distribution which included the Mittag-Leffler distribution as a special case. A random variable x with positive support is said to have a semi–Mittag-Leffler distribution if its Laplace transform is given by

$$\phi(t) = \frac{1}{1+\eta(t)}$$

where $\eta(t)$ satisfies the functional equation $\eta(t) = a\eta(bt)$ where $0 < b < 1$ and α is the unique solution of $ab^\alpha = 1$. It may be denoted by SML(α). Then it follows that $\eta(bt) = b^\alpha h(t)$ where $h(t)$ is a periodic function in t with period $\frac{-\ln b}{2\pi\alpha}$. When $h(t)$ is a constant, the distribution reduces to the Mittag-Leffler distribution. The semi–Mittag-Leffler distribution is also geometrically infinitely divisible and belongs to class L.

2.2.8. α–Laplace distribution

The α–Laplace distribution has characteristic function given by $\varphi(t) = \frac{1}{1+|t|^\alpha}$; $0 < \alpha \le 2$, $-\infty < t < \infty$. This is also called Linnik's distribution. Pillai (1985) refers to it as the α–Laplace distribution since $\alpha = 2$ corresponds to the Laplace distribution. It is unimodal, geometrically strictly stable and belongs to class L. It is normally attracted to the symmetric stable law with exponent α. Also

$$E(|x|^\delta) = \frac{2^\delta \quad \Gamma\left(1+\frac{\delta}{\alpha}\right)\Gamma\left(1-\frac{\delta}{\alpha}\right)\Gamma((1+\delta)/2)}{\sqrt{\pi}\ \ \Gamma\left(1-\frac{\delta}{2}\right)}$$

where $0 < \delta < \alpha$; $0 < \alpha \le 2$.

If u and v are independent random variables where u is exponential with unit mean and v is symmetric stable with exponent α, then $x = u^{1/\alpha}v$ is distributed as α–Laplace. Using this result, Devroye (1990) develops an algorithm for generating random variables having α–Laplace distribution.

Pillai (1985) introduced a larger class of distributions called semi–α–Laplace distribution, with characteristic function given by

$$\varphi(t) = \frac{1}{1 + \eta(t)}$$

where $\eta(t)$ satisfies the functional equation $\eta(t) = a\eta(bt)$ for $0 < b < 1$ and a is the unique solution of $ab^{\alpha} = 1$, $0 < \alpha \leq 2$. Here b is called the order and α is called the exponent of the distribution. If b_1 and b_2 are the orders of the distribution such that $\frac{\ln b_1}{\ln b_2}$ is irrational, then $\eta(t) = c|t|^{\alpha}$, where c is some constant. Pillai (1985) established that, for a semi–α–Laplace distribution with exponent α, $E|x|^{\delta}$ exists for $0 \leq \delta < \alpha$. It can be shown that

$$\varphi(t) = \frac{1}{1 + |t|^{\alpha}[1 - A\cos(k \ln |t|)]}$$

where $k = \frac{2\pi}{\ln b}$, $0 < b < 1$ is the characteristic function of a semi–α–Laplace distribution for suitable choice of A and $\alpha < 1$.

The semi–α–Laplace distribution is also geometrically infinitely divisible and belongs to class L. It is useful in modelling household income data. Mohan *et al.*(1993) refer to it as a geometrically right semi–stable law.

2.2.9. Semi–Pareto distribution

The semi–Pareto distribution was introduced by Pillai (1991). A random variable x with positive support has semi–Pareto distribution $SP(\alpha, p)$ if its survival function is given by $F_x(x) = P(x > x_0) = \frac{1}{1+\psi(x_0)}$ where $\psi(x_0)$ satisfies the functional equation $p\psi(x) = \psi(p^{1/\alpha}x)$; $0 < p < 1,\ \alpha > 0$.

The above definition is analogous to that of the semi–stable law defined by Levy (see Pillai (1971)). It can be shown that $\psi(x) = x^{\alpha}h(x)$ where $h(x)$ is periodic in $\ln x$ with period $\frac{-2\pi\alpha}{\ln p}$. For example if $h(x) = \exp[\beta\cos(\alpha \ln x)]$, then it satisfies the above functional equation with $p = \exp(-2\pi)$ and $\psi(x)$ monotone

increasing with $0 < \beta < 1$. The semi–Pareto distribution can be viewed as a more general class which includes the Pareto type III distribution when $\psi(x) = cx^{\alpha}$, where c is a constant.

Exercises 2.2.

2.2.1. Examine whether the following distributions are infinitely divisible.

(i) normal (ii) exponential (iii) Laplace (iv) Cauchy
(v) binomial (vi) Poisson (vii) Geometric (viii) negative binomial

2.2.2. Show that exponential distribution is geometric infinite divisible and self decomposable.

2.2.3. Examine whether Cauchy distribution is self-decomposable.

2.2.4. Show that (i) Mittag-Leffler distribution is g.i.d and belongs to class L. (ii) α-Laplace distribution is g.i.d and self-decomposable.

2.2.5. Give a distribution which is infinitely divisible but not g.i.d.

2.2.6. Show that AR(1) structure $x_n = ax_{n-1} + \epsilon_n; a \in (0, 1)$ is stationary Markovian if and only if $\{x_n\}$ is self decomposable.

2.2.7. Show that geometric and negative binomial distribution are discrete self-decomposable.

2.2.8. Consider the symmetric stable distribution with characteristic function $\varphi(t) = \mathrm{e}^{-|t|^{\alpha}}$. Is it self-decomposable?

2.3. Stationary Time Series Models

2.3.1. Introduction

A time series is a realization of a stochastic process. In other words, a time series, $\{x_t\}$, is a family of real–valued random variables indexed by $t \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integers. More specifically, it is referred to as a discrete parameter time series. The time series $\{x_t\}$ is said to be stationary if, for any $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, any $k \in \mathbb{Z}$, and $n = 1, 2, \ldots,$

$$F_{x_{t_1}, x_{t_2}, \ldots, x_{t_n}}(x_1, x_2, \ldots, x_n) = F_{x_{t_1+k}, x_{t_2+k}, \ldots, x_{t_n+k}}(x_1, x_2, \ldots, x_n)$$

where F denotes the distribution function of the set of random variables which appear as suffices. This is called stationarity in the strict sense.

Less stringently, we say a process $\{x_n\}$ is weakly stationary if the mean and variance of x_t remain constant over time and the covariance between any two values x_t and x_s depends only on the time difference and not on their individual time points.

$\{x_t\}$ is called a Gaussian process if, for all $t_n; n \geq 1$ the set of random variables $\{x_{t_1}, x_{t_2}, \ldots, x_{t_n}\}$ has a multivariate normal distribution.

Since a multivariate normal distribution is completely specified by its mean vector and covariance matrix, it follows that for a Gaussian process weak stationarity implies complete stationarity. But for non–Gaussian processes, this may not hold.

2.3.2. Autoregressive models

The era of linear time series models began with autoregressive models first introduced by Yule in 1927. The standard form of an autoregressive model of order p, denoted by AR(p), is given by

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + \epsilon_t; \qquad t = 0, \pm 1, \pm 2, \ldots$$

where $\{\epsilon_t\}$ are independent and identically distributed random variables called innovations and a_j, p are fixed parameters, with $a_p \neq 0$.

Another kind of model of great practical importance in the representation of observed time series is the moving average model. The standard form of a moving average model of order q, denoted by MA(q), is given by $x_t = \sum_{j=1}^{q} b_j \epsilon_{t-j} + \epsilon_t$; $t \in \mathbb{Z}$ where b_j, q are fixed parameters, with $b_q \neq 0$.

To achieve greater flexibility in the fitting of actually observed time series, it is more advantageous to include both autoregressive and moving average terms in the model. Such models called autoregressive–moving average models, denoted by ARMA (p,q), have the form

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + \sum_{k=1}^{q} b_k \epsilon_{t-k} + \epsilon_t; \qquad t \in \mathbb{Z}$$

where $\{a_j\}_{j=1}^{p}$ and $\{b_k\}_{k=1}^{q}$ are real constants called parameters of the model. It can be seen that an AR(p) model is the same as an ARMA(p,o) model and a MA(q) model is the same as an ARMA(o,q) model.

With the introduction of various non–Gaussian and non–linear models, the standard form of autoregression was widened in several respects.

A more general definition of autoregression of order p is given in terms of the linear conditional expectation requirement that

$$E(x_t | x_{t-1}, x_{t-2}, \ldots) = \sum_{j=1}^{p} a_j x_{t-j}$$

This definition could apply to models which are not of the linear form (see Lawrance (1991)).

2.3.3. A general solution

We consider a first order autoregressive model with innovation given by the structural relationship

$$x_n = \epsilon_n + \begin{cases} 0 & \text{with probability } p \\ x_{n-1} & \text{with probability } 1-p \end{cases} \tag{2.3.1}$$

where $p \in (0, 1)$ and $\{\epsilon_n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables selected in such a way that $\{x_n\}$ is stationary Markovian with a given marginal distribution function F.

Let $\phi_x(t) = E[e^{-tx}]$ be the Laplace–Stieltjes transform of x. Then (2.3.1) gives

$$\phi_{x_n}(t) = \phi_{\epsilon_n}(t)[p + (1-p)\phi_{x_{n-1}}(t)]$$

If we assume stationarity, this simplifies to

$$\phi_\epsilon(t) = \frac{\phi_x(t)}{p + (1-p)\phi_x(t)} \tag{2.3.2}$$

or equivalently

$$\phi_x(t) = \frac{p\phi_\epsilon(t)}{1 - (1-p)\phi_\epsilon(t)}. \tag{2.3.3}$$

When $\{x_n\}$ is marginally distributed as exponential, it is easy to see that (2.3.1) gives the TEAR(1) model.

We note that $\phi_\epsilon(t)$ in (2.3.2) does not represent a Laplace transform always. In order that the process given by (2.3.1) is properly defined, there should exist an innovation distribution such that $\phi_\epsilon(t)$ is a Laplace transform for all $p \in (0,1)$. To establish the main results we need the following lemmas.

(Pillai (1990b)) Let F be a distribution with positive support and $\phi(t)$ be its Laplace transform. Then F is geometrically infinitely divisible if and only if

$$\phi(t) = \frac{1}{1 + \psi(t)}$$

where $\psi(t)$ is Bernstein with $\psi(0) = 0$.

Now we consider the following definition from Pillai (1990b).

Definition 2.3.1. For any non–vanishing Laplace transform $\phi(t)$, the function
$\psi(t) = \dfrac{1}{\phi(t)} - 1$ is called the third characteristic.

Lemma 2.3.1. Let $\psi(t)$ be the third characteristic of $\phi(t)$. Then $p\psi(t)$ is a third characteristic for all $p \in (0,1)$ if and only if $\psi(t)$ has complete monotone derivative and $\psi(0) = 0$.

Thus we have the following theorem.

Theorem 2.3.1. *$\phi_\epsilon(t)$ in* (2.3.2) *represents a Laplace transform for all $p \in (0,1)$ if and only if $\phi_x(t)$ is the Laplace transform of a geometrically infinitely divisible distribution.*

This leads to the following theorem which brings out the role of geometrically infinitely divisible distributions in defining the new first order autoregressive model given by (2.3.1).

Theorem 2.3.2. *The innovation sequence $\{\epsilon_n\}$ defining the first order autoregressive model given by*

$$x_n = \epsilon_n + \begin{cases} 0 & \text{with probability } p \\ x_{n-1} & \text{with probability } 1-p \end{cases}$$

where $p \in (0,1)$, exists if and only if the stationary marginal distribution of x_n is geometrically infinitely divisible. Then the innovation distribution is also geometrically infinitely divisible.

Proof 2.3.1. Suppose that an innovation sequence $\{\epsilon_n\}$ such that the model (2.3.1) is properly defined exists. This implies that $\phi_\epsilon(t)$ in (2.3.3) is a Laplace transform for all $p \in (0,1)$. Then from (2.3.3)

$$\begin{aligned} \phi_x(t) &= p\phi_\epsilon(t)[1-(1-p)\phi_\epsilon(t)]^{-1} \\ &= \sum_{n=1}^{\infty} p(1-p)^{n-1}[\phi_\epsilon(t)]^n \end{aligned}$$

showing that the stationary marginal distribution of x_n is geometrically infinitely divisible. Conversely, if x_n has a stationary marginal distribution which is geometrically infinitely divisible, then $\phi_x(t) = \dfrac{1}{1+\psi(t)}$ where $\psi(t)$ has complete monotone derivative and $\psi(0) = 0$. Then from (2.3.2) we get $\phi_\epsilon(t) = \dfrac{1}{1+p\psi(t)}$, which establishes the existence of an innovation distribution, which is geometrically infinitely divisible.

2.3.4. Extension to a k-th order autoregressive model

In this section we consider an extension of the model given by (2.3.1) to the k-th order. The structure of this model is given by

$$x_n = \epsilon_n + \begin{cases} 0 & \text{with probability } p_0 \\ x_{n-1} & \text{with probability } p_1 \\ \vdots & \\ x_{n-k} & \text{with probability } p_k \end{cases} \tag{2.3.4}$$

where $p_i \in (0,1)$ for $i = 0, 1, \ldots, k$ and $p_0 + p_1 + \cdots + p_k = 1$. Taking Laplace transforms on both sides of (2.3.4) we get

$$\phi_{x_n}(t) = \phi_{\epsilon_n}(t)\left[p_0 + \sum_{i=1}^{k} p_i \phi_{x_{n-i}}(t)\right]$$

Assuming stationarity, it simplifies to

$$\begin{aligned} \phi_x(t) &= \phi_\epsilon(t)\left[p_0 + \sum_{i=1}^{k} p_i \phi_x(t)\right] \\ &= \phi_\epsilon(t)[p_0 + (1-p_0)\phi_x(t)]. \end{aligned}$$

This yields

$$\phi_\epsilon(t) = \frac{\phi_x(t)}{p_0 + (1-p_0)\phi_x(t)} \tag{2.3.5}$$

which is analogous to the expression (2.3.2).

It may be noted that $k = 1$ corresponds to the first order model with $p = p_0$. From (2.3.5) it follows that the results obtained in Section 2.1.2 hold good for the k-th order model given by (2.3.4). This establishes the importance of geometrically infinitely divisible distributions in autoregressive modelling.

2.3.5. Mittag-Leffler autoregressive structure

The Mittag-Leffler distribution was introduced by Pillai (1990a) and has Laplace transform $\phi(t) = \dfrac{1}{1+t^\alpha}$, $0 < \alpha \leq 1$. When $\alpha = 1$, this corresponds to the exponential distribution with unit mean. Jayakumar and Pillai (1993) considered

the semi–Mittag-Leffler distribution with exponent α. Its Laplace transform is of the form $\frac{1}{1+\eta(t)}$ where $\eta(t)$ satisfies the functional equation

$$\eta(t) = a\eta(bt), \qquad 0 < b < 1 \tag{2.3.6}$$

and a is the unique solution of $ab^{\alpha} = 1$ where $0 < \alpha \leq 1$. Then by Lemma 2.3.1 of Jayakumar and Pillai (1993), the solution of the functional equation (2.3.6) is $\eta(t) = t^{\alpha}h(t)$ where $h(t)$ is periodic in $\ln t$ with period $-\frac{2\pi\alpha}{\ln b}$. When $h(t) = 1$, $\eta(t) = t^{\alpha}$ and hence the Mittag-Leffler distribution is a special case of the semi-Mittag-Leffler distribution. It is obvious that the semi–Mittag-Leffler distribution is geometrically infinitely divisible.

Now we bring out the importance of the semi-Mittag-Leffler distribution in the context of the new autoregressive structure given by (2.3.1). The following theorem establishes this.

Theorem 2.3.3. *For a positive valued first order autoregressive process $\{x_n\}$ satisfying (2.3.1) the stationary marginal distribution of x_n and ϵ_n are identical except for a scale change if and only if x_n's are marginally distributed as semi-Mittag-Leffler.*

Proof 2.3.2. Suppose that the stationary marginal distributions of x_n and ϵ_n are identical. This implies $\phi_{\epsilon}(t) = \phi_x(ct)$ where c is a constant. Then from (2.3.2) we get

$$\phi_x(ct) = \frac{\phi_x(t)}{p + (1-p)\phi_x(t)} \tag{2.3.7}$$

Writing $\phi_x(t) = \frac{1}{1+\eta(t)}$ in (2.1.7) we get

$$\frac{1}{1+\eta(ct)} = \frac{1}{1+p\eta(t)}$$

so that

$$\eta(ct) = p\eta(t)$$

By choosing $c = p^{1/\alpha}$, it follows that x_n is distributed as semi–Mittag-Leffler with exponent α.

Conversely, we assume that the stationary marginal distribution of x_n is semi-Mittag-Leffler. Then from (2.3.2)

$$\phi_\epsilon(t) = \frac{1}{1 + p\eta(t)} = \frac{1}{1 + \eta(p^{1/\alpha}t)}.$$

This establishes that $\epsilon_n \stackrel{d}{=} p^{1/\alpha} M_n$ where $\{M_n\}$ are independently and identically distributed as semi–Mittag-Leffler.

It can be easily seen that the above result is true in the case of the k-th order autoregressive model given by (2.3.4) also.

Exercises 2.3.

2.3.1. Define an AR(1) process and obtain the stationary solution for the distribution of $\{\epsilon_n\}$ when $\{x_n\}$ are exponentially distributed.

2.3.2. Show that an AR(1) model can be expressed as $MA(\infty)$ model.

2.3.3. Consider a new AR(1) model with exponential innovations.

2.3.4. Examine whether two-parameter Gamma distribution is g.i.d., giving conditions if any.

2.3.5. Show that exponential distribution is a special case of Mittag-Leffler distribution.

2.3.6. Obtain the stationary distribution of $\{\epsilon_n\}$ in the AR(1) structure $x_n = ax_{n-1} + \epsilon_n; a \in (0, 1)$ when $\{x_n\}$ follows exponential distribution. Generalize it to the case of Mittag-Leffler random variables.

2.3.7. Obtain the structure of the innovation distribution if $\{x_n\}$ follows α-Laplace distribution where $x_n = ax_{n-1} + \epsilon_n$. Deduce the case when $\alpha = 2$.

2.3.8. Show that if $\{x_n\}$ follows Cauchy distribution then $\{\epsilon_n\}$ also follows a Cauchy distribution in the AR(1) equation $x_n = ax_{n-1} + \epsilon_n$.

2.4. A Structural Relationship and New Processes

In this section we obtain the specific structural relationship between the stationary marginal distributions of x_n and ϵ_n in the new autoregressive model.

Fujita (1993) generalized the results on Mittag-Leffler distributions and obtained a new characterization of geometrically infinitely divisible distributions with positive support using Bernstein functions. It was established that a distribution function G with $G(0) = 0$ is geometrically infinitely divisible if and only if G can be expressed in the form.

$$G(x) = \sum_{n=1}^{\infty}(-1)^{n+1}\lambda^n W^{n*}([0,x]); \qquad x > 0,\ \lambda > 0 \tag{2.4.1}$$

where $W^{n*}(dx)$ is the n-fold convolution measure of a unique positive measure $W(dx)$ on $[0,\infty)$ such that

$$\frac{1}{f(x)} = \int_0^{\infty} e^{-sx}W(ds); \qquad x > 0 \tag{2.4.2}$$

for some Bernstein function f such that $\lim_{x\downarrow o}(x) = 0$ and $\lim_{x\to\infty} f(x) = \infty$. Then the Laplace transform of $G(x)$ is $\dfrac{\lambda}{\lambda + f(t)}$. Using this result we get the following theorem.

Theorem 2.4.1. *The k-th order autoregressive equation given by (2.3.4) defines a stationary process with a given marginal distribution function $F_x(x)$ for x_n if and only if $F_x(x)$ can be expressed in the form*

$$F_x(x) = \sum_{n=1}^{\infty}(-1)^{n+1}\lambda^n W^{n*}([0,x]); \qquad x > 0,\ \lambda > 0. \tag{2.4.3}$$

Then the innovations $\{\epsilon_n\}$ have a distribution function $F_\epsilon(x)$ given by

$$F_\epsilon(x) = \sum_{n=1}^{\infty}(-1)^{n+1}(\lambda/p_0)^n W^{n*}([0,x]); \qquad x > 0,\ \lambda > 0, \tag{2.4.4}$$

where $p_0 \in (0,1)$ and W^{n*} is as in (2.4.1).

Proof 2.4.1. We have from Theorem 2.3.1 that $F_x(x)$ is geometrically infinitely divisible. Then (2.4.3) follows directly from Fujita (1993).

Now by substituting $\phi_x(t) = \dfrac{\lambda}{\lambda + f(t)}$ in (2.1.2) we get

$$\phi_\epsilon(t) = \frac{\lambda}{\lambda + p_0 f(t)} = \frac{(\lambda/p_0)}{(\lambda/p_0) + f(t)}$$

which leads to (2.4.4). This completes the proof.

The above theorem can be used to construct various autoregressive models under different stationary marginal distributions for x_n.

For example, the TEAR(1) model of Lawrance and Lewis (1981) can be obtained by taking $f(t) = t$. Then $W^{n*}([0, x]) = \dfrac{x^n}{n!}$ so that $F_x(x) = 1 - \mathrm{e}^{-\lambda x}$ and $F_\epsilon(x) = 1 - \mathrm{e}^{-(\lambda/p)x}$. If we take $\lambda = 1$ and $f(t) = t^\alpha$; $0 < \alpha \leq 1$ we can obtain an easily tractable first order autoregressive Mittag-Leffler process denoted by TMLAR(1). In this case $W^{n*}([0, x]) = \dfrac{x^{n\alpha}}{\Gamma(1 + n\alpha)}$. In a similar manner by taking $\lambda = 1$ and $f(t)$ satisfying the functional equation $f(t) = af(bt)$ where $a = b^{-\alpha}$; $0 < b < 1$, $0 < \alpha \leq 1$, we can obtain an easily tractable first order autoregressive semi–Mittag-Leffler process denoted by TSMLAR(1).

2.4.1. The TMLAR(1) process

An easily tractable form of a first order autoregressive Mittag-Leffler process, called TMLAR(1), is constituted by $\{x_n\}$ having a structure of the form

$$x_n = p^{1/\alpha} M_n + \begin{cases} 0 & \text{with probability p} \\ x_{n-1} & \text{with probability 1-p} \end{cases} \tag{2.4.5}$$

where $p \in (0, 1)$; $0 < \alpha \leq 1$ and $\{M_n\}$ is independently and identically distributed as Mittag-Leffler with exponent α and $x_0 \stackrel{d}{=} M_1$. The model (2.4.5) can be rewritten in the form

$$x_n = p^{1/\alpha} M_n + I_n x_{n-1} \tag{2.4.6}$$

where $\{I_n\}$ is a Bernoulli sequence such that $P(I_n = 0) = p$ and $P(I_n = 1) = 1 - p$.

If in the structural form (2.4.5), we assume that $\{M_n\}$ are distributed as semi–Mittag-Leffler with exponent α, then $\{x_n\}$ constitute a tractable semi–Mittag-Leffler autoregressive process of order 1, called TSMLAR(1). Both models are Markovian and stationary. It can be seen that the TMLAR(1) process is a special case of the TSMLAR(1) process since the Mittag-Leffler distribution is a special case of the semi–Mittag-Leffler distribution.

Now we shall consider the TSMLAR(1) process and establish that it is strictly stationary and Markovian, provided x_0 is distributed as semi–Mittag-Leffler. In order to prove this we use the method of induction.

Suppose that x_{n-1} is distributed as semi–Mittag-Leffler (α). Then by taking Laplace transforms on both sides of (2.4.5), we get

$$\begin{aligned}\phi_{x_n}(t) &= \phi_{M_n}(p^{1/\alpha}t)[p + (1-p)\phi_{x_{n-1}}(t)] \\ &= \frac{1}{1+\eta(p^{1/\alpha}t)}\left[p + (1-p)\frac{1}{1+\eta(t)}\right] \\ &= \frac{1}{1+p\eta(t)} \cdot \frac{1+p\eta(t)}{1+\eta(t)} \\ &= \frac{1}{1+\eta(t)}.\end{aligned}$$

Hence x_n is distributed as semi–Mittag-Leffler with exponent α.

If x_0 is arbitrary, then also it is easy to establish that $\{x_n\}$ is asymptotically stationary. Thus we have the following theorem.

Theorem 2.4.2. *The first order autoregressive equation*

$$x_n = p^{1/\alpha}M_n + I_n x_{n-1}; \qquad n = 1, 2, \ldots, \quad p \in (0,1)$$

where $\{I_n\}$ *are independent Bernoulli random variables such that* $P(I_n = 0) = p = 1 - P(I_n = 1)$ *defines a positive valued strictly stationary first order autoregressive process if and only if* $\{M_n\}$ *are independently and identically distributed as semi–Mittag-Leffler with exponent* α *and* $x_0 \stackrel{d}{=} M_1$.

Remark 2.4.1. If we consider characteristic functions instead of Laplace transforms, the results can be applied to real valued autoregressive processes. Then

the role of semi–Mittag-Leffler distributions is played by semi–α–Laplace distributions introduced by Pillai (1985).

2.4.2. The NEAR(1) model

In this section we consider a generalized form of the first order autoregressive equation. The new structure is given by

$$x_n = \epsilon_n + \begin{cases} 0 & \text{with probability } p \\ ax_{n-1} & \text{with probability } 1-p \end{cases} \tag{2.4.7}$$

where $0 \le p \le 1$; $0 \le a \le 1$ and $\{\epsilon_n\}$ is a sequence of independent and identically distributed random variables such that $\{x_n\}$ have a given stationary marginal distribution. Let $\phi_x(t) = E[e^{-tx}]$ be the Laplace–Stieltjes transform of x. Then (2.4.7) gives

$$\phi_{x_n}(t) = \phi_{\epsilon_n}(t)[p + (1-p)\phi_{x_{n-1}}(at)]$$

Assuming stationarity, it simplifies to

$$\phi_\epsilon(t) = \frac{\phi_x(t)}{p + (1-p)\phi_x(at)}. \tag{2.4.8}$$

When $p = 0$ and $0 < a < 1$, the model (2.4.7) is the standard first order autoregressive model. Then the model is properly defined if and only if the stationary marginal distribution of x_n is self–decomposable. When $a = 1$, $0 < p < 1$ the model is the same as the model (2.4.1), which is properly defined if and only if the stationary marginal distribution of x_n is geometrically infinitely divisible. When $a = 0$ or $p = 1$, x_n and ϵ_n are identically distributed.

Now we consider the case when $a \in (0, 1]$ and $p \in (0, 1]$, but not simultaneously equal to 1. Lawrance and Lewis (1981) developed an NEAR(1) model with exponential (λ) marginal distribution for x_n. Then $\phi_x(t) = \dfrac{\lambda}{\lambda + t}$ and substitution in (2.4.8) gives

$$\phi_\epsilon(t) = \frac{\lambda + at}{\lambda + t} \cdot \frac{\lambda}{\lambda + pat} \tag{2.4.9}$$

which can be rewritten as

$$\phi_\epsilon(t) = \left(\frac{1-a}{1-pa}\right)\left(\frac{\lambda}{\lambda+t}\right) + \left[\frac{(1-p)a}{1-pa}\right]\left(\frac{\lambda}{\lambda+pat}\right).$$

Hence ϵ_n can be regarded as a convex exponential mixture of the form

$$\epsilon_n = \begin{cases} E_n & \text{with probability } \frac{1-a}{1-pa} \\ paE_n & \text{with probability } \frac{(1-p)a}{1-pa} \end{cases} \tag{2.4.10}$$

where $\{E_n\}$; $n = 1, 2, \ldots$ are independent and identically distributed as exponential (λ) random variables. Another representation for ϵ_n can be obtained from (2.4.9) by writing

$$\phi_\epsilon(t) = \left[a + (1-a)\frac{\lambda}{\lambda+t}\right]\left[\frac{\lambda}{\lambda+pat}\right]. \tag{2.4.11}$$

Then writing w.p. for 'with probability' ϵ_n can be regarded as the sum of two independent random variables u_n and v_n where

$$u_n = \begin{cases} 0 & \text{w. p. } a \\ E_n & \text{w. p. } 1-a \text{ and} \end{cases} \tag{2.4.12}$$

$$v_n = paE_n$$

where $\{E_n\}$; $n = 1, 2, \ldots$ are exponential (λ). It may be noted that when $p = 0$, the model is identical with the EAR(1) process, of Gaver and Lewis (1980). Thus the new representation of ϵ_n seems to be more appropriate, when NEAR(1) process is regarded as a generalization of the EAR(1) process.

2.4.3. New Mittag-Leffler autoregressive models

Now we construct a new first order autoregressive process with Mittag-Leffler marginal distribution, called the NMLAR(1) model.

The structure of the model is as in (2.4.7) and the innovations can be derived by substituting $\phi_x(t) = \dfrac{1}{1+t^\alpha}$; $0 < \alpha \leq 1$ in (2.4.8). This gives

$$\phi_\epsilon(t) = \frac{1+a^\alpha t^\alpha}{1+t^\alpha} \cdot \frac{1}{1+a^\alpha pt^\alpha}.$$

Hence the innovations ϵ_n can be given in the form

$$\epsilon_n = \begin{cases} M_n & \text{with probability } \frac{1-a^\alpha}{1-pa^\alpha} \\ pa^\alpha M_n & \text{with probability } \frac{(1-p)a^\alpha}{1-pa^\alpha} \end{cases} \tag{2.4.13}$$

where $\{M_n\}$ are Mittag-Leffler (α) random variables.

An alternate representation of ϵ_n is $\epsilon_n = u_n + v_n$ where u_n and v_n are independent random variables such that

$$u_n = \begin{cases} 0 & \text{w.p. } a^\alpha \\ M_n & \text{w.p. } 1 - a^\alpha \text{ and} \end{cases} \tag{2.4.14}$$

$$v_n = ap^{1/\alpha} M_n$$

where $\{M_n\}$; $n = 1, 2, \ldots$ are independent Mittag-Leffler (α) random variables.

It can be shown that the process is strictly stationary and Markovian. This gives us the following theorem.

Theorem 2.4.3. *The first order autoregressive equation given by* (2.4.7) *defines a strictly stationary* AR(1) *process with a Mittag-Leffler* (α) *marginal distribution for* x_n *if and only if the innovations are of the form* $\epsilon_n = u_n + v_n$ *where* u_n *and* v_n *are as in* (2.4.14) *with* x_0 *distributed as Mittag-Leffler* (α).

Proof 2.4.2. We prove this by induction. We assume that x_{n-1} is Mittag-Leffler (α). Then by taking Laplace transforms, we get

$$\begin{aligned} \phi_{x_n}(t) &= \phi_{M_n}(ap^{1/\alpha}t) \cdot [a^\alpha + (1 - a^\alpha)\phi_{M_n}(t)] \\ &\quad \times [p + (1 - p)\phi_{x_{n-1}}(at)] \\ &= \frac{1}{1 + a^\alpha pt^\alpha} \cdot \left[a^\alpha + (1 - a^\alpha)\frac{1}{1 + t^\alpha}\right] \\ &\quad \times \left[p + (1 - p)\frac{1}{1 + a^\alpha t^\alpha}\right] \\ &= \frac{1}{1 + a^\alpha pt^\alpha} \cdot \frac{1 + a^\alpha t^\alpha}{1 + t^\alpha} \cdot \frac{1 + pa^\alpha t^\alpha}{1 + a^\alpha t^\alpha} \\ &= \frac{1}{1 + t^\alpha}. \end{aligned}$$

This shows that x_n is distributed as Mittag-Leffler (α), and this establishes the sufficiency part.

The necessary part is obvious from the derivation of the innovation sequence. This completes the proof.

The joint distribution of (x_n, x_{n-1}) is of interest in describing the process and matching it with data. Therefore, we shall obtain the joint distribution with the use of Laplace-Stieltjes transforms. The bivariate Laplace transform is given by

$$\begin{aligned}\phi_{x_n,x_{n-1}}(s,t) &= E\{\exp(-sx_n - tx_{n-1})\} \\ &= \phi_\epsilon(s)\{p\phi_x(t) + (1-p)\phi_x(as+t)\} \\ &= \frac{1+a^\alpha s^\alpha}{1+s^\alpha} \cdot \frac{1}{1+pa^\alpha s^\alpha} \cdot \left\{ \frac{p}{1+t^\alpha} + \frac{1-p}{1+(as+t)^\alpha} \right\}.\end{aligned}$$

It is possible to obtain the joint distribution by inverting this expression.

2.4.4. The NSMLAR(1) process

Now we extend the NMLAR(1) process to a wider class to construct a new semi–Mittag-Leffler first order autoregressive process. The process has the structure

$$x_n = \epsilon_n + \begin{cases} 0 & \text{with probability } p \\ ax_{n-1} & \text{with probability } 1-p \end{cases}$$

where $\{\epsilon_n\}$ are independently and identically distributed as the sum of two independent random variables u_n and v_n where

$$\begin{aligned} u_n &= \begin{cases} 0 & \text{w.p. } a^\alpha \\ M_n & \text{w.p. } 1-a^\alpha \text{ and} \end{cases} \\ v_n &= ap^{1/\alpha} M_n \end{aligned} \tag{2.4.15}$$

where $\{M_n\}$; $n = 1, 2, \ldots$ are independently and identically distributed as semi–Mittag-Leffler (α).

This process is also clearly strictly stationary and Markovian provided x_0 is semi–Mittag-Leffler (α). This follows by induction. In terms of Laplace transforms we have

$$\begin{aligned}
\phi_{x_n}(t) &= [p + (1-p)\phi_{x_{n-1}}(at)][a^\alpha + (1-a^\alpha)\phi_{M_n}(t)] \\
&\quad \times [\phi_{M_n}(ap^{1/\alpha}t)] \\
&= \left[p + (1-p)\cdot\frac{1}{1+\eta(at)}\right]\left[a^\alpha + (1-a^\alpha)\cdot\frac{1}{1+\eta(t)}\right] \\
&\quad \times \frac{1}{1+\eta(ap^{1/\alpha}t)} \\
&= \left[p + (1-p)\frac{1}{1+a^\alpha\eta(t)}\right]\cdot\left[\frac{1+a^\alpha\eta(t)}{1+\eta(t)}\right] \\
&\quad \times \frac{1}{1+a^\alpha p\eta(t)} \\
&= \frac{1}{1+\eta(t)}.
\end{aligned}$$

Thus we have established the following theorem.

Theorem 2.4.4. *The first order autoregressive equation*

$$x_n = aI_n x_{n-1} + \epsilon_n; \quad n = 1, 2, \ldots$$

where $\{I_n\}$ *are independent Bernoulli sequences such that* $P(I_n = 0) = p$ *and* $P(I_n = 1) = 1 - p$; $p \in (0,1)$, $a \in (0,1)$ *is a strictly stationary AR(1) process with semi–Mittag-Leffler* (α) *marginal distribution if and only if* $\{\epsilon_n\}$ *are independently and identically distributed as the sum of two independent random variables* u_n *and* v_n *as in* (2.4.15) *and* x_0 *is distributed as semi–Mittag-Leffler* (α).

When $\eta(t) = t^\alpha$, the NSMLAR(1) model becomes the NMLAR(1) model.

Remark 2.4.2. If we consider characteristic functions instead of Laplace transforms, the results can be applied to real valued autoregressive processes. Then the role of semi-Mittag-Leffler distributions is played by semi-α-Laplace distributions introduced by Pillai (1985). As special cases we get Laplace and α-Laplace processes.

Exercises 2.4.

2.4.1. If $f(t) = t$, find $W^{n*}([0, x])$.

2.4.2. If $f(t) = t^{\alpha}$, find $F_{\epsilon}(x)$.

2.4.3. State any three distributions belonging to the semi-Mittag-Leffler family.

2.4.4. Show that the stationary solution of Equation 4.5.7 is a family consisiting of g.i.d. and class L distributions.

2.4.5. Obtain the innovation structure of the NEAR(1) model.

2.4.6. Obtain the innovation structure of the NMLAR(1) model.

2.5. Tailed Processes

In an attempt to develop autoregressive models for time series with exact zeroes Littlejohn (1993) formulated an autoregressive process with exponential tailed marginal distribution, after the new exponential autoregressive process (NEAR(1)) of Lawrance and Lewis (1981). However, the primary aim of Littlejohn was to extend the time reversibility theorem of Chernick *et al.*(1988) and hence the model was not studied in detail. Hence we intend to make a detailed study on this process. Here the tail of a non–negative random variable refers to the positive part of the sample space, excluding only the point zero.

Definition 2.5.1. A random variable E is said to have the exponential tailed distribution denoted by $ET(\lambda, \theta)$ if $P(E = 0) = \theta$ and $P(E > x) = (1 - \theta)\mathrm{e}^{-\lambda x}$; $x > 0$ where $\lambda > 0$ and $0 \leq \theta < 1$. Then the Laplace-Stieltjes transform of E is given by

$$\begin{aligned}\phi_E(t) &= \theta + (1 - \theta)\frac{\lambda}{\lambda + t} \\ &= \frac{\lambda + \theta t}{\lambda + t}\end{aligned}$$

2.5.1. The exponential tailed autoregressive process [ETAR(1)]

It is evident that the exponential tailed distribution is not self–decomposable and so it cannot be marginal to the autoregressive structure of Gaver and Lewis (1980). But an autoregressive process satisfying the NEAR(1) structure given by (2.4.7) can be constructed as follows.

We have from (2.4.8), by substituting $\phi_x(t) = \dfrac{\lambda + \theta t}{\lambda + t}$, the Laplace transform of the innovation ϵ_n in the stationary case as

$$\begin{aligned}\phi_\epsilon(t) &= \left[\frac{\lambda + \theta t}{\lambda + t}\right]\left[\frac{\lambda + at}{\lambda + a[p + (1-p)\theta]t}\right] \\ &= \left[\frac{\lambda + at}{\lambda + t}\right]\left[\frac{\lambda + \theta t}{\lambda + bt}\right]\end{aligned}$$

where $b = a[p + (1-p)\theta]$

$$\phi_\epsilon(t) = \left[a + (1-a)\frac{\lambda}{\lambda + t}\right]\left[\frac{\theta}{b} + \left(1 - \frac{\theta}{b}\right)\frac{(\lambda/b)}{(\lambda/b) + t}\right]$$

so that the innovations $\{\epsilon_n\}$ can be represented as the sum of two independent exponential tailed random variables u_n and v_n where

$$u_n \stackrel{d}{=} ET(\lambda, a) \qquad \text{and} \quad v_n \stackrel{d}{=} ET(\lambda', \theta') \tag{2.5.1}$$

where $\lambda' = \lambda/b$ and $\theta' = \theta/b$, provided $\theta \leq b$. Since $p \leq 1$, we require that $\theta \leq a$. Thus the ETAR(1) process can be defined as a sequence $\{x_n\}$ satisfying (2.4.7) where $\{\epsilon_n\}$ is a sequence of independent and identically distributed random variables such that $\epsilon_n = u_n + v_n$ where u_n and v_n are as in (2.5.1).

It can be easily shown that the process is strictly stationary and Markovian provided x_0 is distributed as $ET(\lambda, \theta)$. This follows by mathematical induction since

$$\begin{aligned}\phi_{x_n}(t) &= \phi_{\epsilon_n}(t)\cdot[p+(1-p)\phi_{x_{n-1}}(at)]\\ &= \frac{\lambda+at}{\lambda+t}\cdot\frac{\lambda+\theta t}{\lambda+bt}\cdot\left[p+(1-p)\left(\frac{\lambda+\theta at}{\lambda+at}\right)\right]\\ &= \frac{\lambda+at}{\lambda+t}\cdot\frac{\lambda+\theta t}{\lambda+bt}\cdot\frac{\lambda+bt}{\lambda+at}\\ &= \frac{\lambda+\theta t}{\lambda+t}.\end{aligned}$$

When $\theta = 0$, the $ET(\lambda,\theta)$ distribution reduces to the exponential (λ) distribution and the ETAR(1) model then becomes the NEAR(1) model.

2.5.2. The Mittag-Leffler tailed autoregressive process [ML-TAR(1)]

The Mittag-Leffler tailed distribution has Laplace transform given by

$$\begin{aligned}\phi_x(t) &= \theta+(1-\theta)\cdot\frac{1}{1+t^\alpha}\\ &= \frac{1+\theta t^\alpha}{1+t^\alpha}; \qquad 0<\alpha\le 1\end{aligned}$$

and the distribution shall be denoted by MLT (α,θ). Similarly for a two-parameter Mittag-Leffler random variable ML(α,λ) the Laplace transform of the tailed Mittag-Leffler distribution is given by $\phi_x(t) = \theta+(1-\theta)\frac{\lambda^\alpha}{\lambda^\alpha+t^\alpha} = \frac{\lambda^\alpha+\theta t^\alpha}{\lambda^\alpha+t^\alpha}$. This shall be denoted by MLT(α,λ,θ). The MLTAR(1) process has the general structure given by the equation (2.4.7). The innovation structure can be derived as follows.

$$\begin{aligned}\phi_\epsilon(t) &= \frac{1+\theta t^\alpha}{1+t^\alpha}\cdot\frac{1+a^\alpha t^\alpha}{1+a^\alpha[p+(1-p)\theta]t^\alpha}\\ &= \frac{1+a^\alpha t^\alpha}{1+t^\alpha}\cdot\frac{1+\theta t^\alpha}{1+ct^\alpha}\end{aligned}$$

where $c = a^\alpha[p+(1-p)\theta]$. Therefore

$$\phi_\epsilon(t) = \left[a^\alpha+(1-a^\alpha)\frac{1}{1+t^\alpha}\right]\left[\frac{\frac{1}{c}+\frac{\theta}{c}t^\alpha}{\frac{1}{c}+t^\alpha}\right].$$

Hence the innovation $\{\epsilon_n\}$ can be viewed as the sum of two independently distributed random variables u_n and v_n where

$$u_n \stackrel{d}{=} MLT(\alpha, a^\alpha)$$

and

$$v_n \stackrel{d}{=} MLT\,(\alpha, \lambda', \theta')$$

where $\lambda' = 1/c^{1/\alpha}$ and $\theta' = \theta/c$ provided $\theta \leq c$. This holds when $\theta \leq a^\alpha$.

The model can be extended to the class of semi–Mittag-Leffler distributions. Here we consider a semi–Mittag-Leffler distribution with Laplace transform

$$\phi_x(t) = \frac{\lambda^\alpha}{\lambda^\alpha + \eta(t)}$$

where $\eta(t)$ satisfies the functional equation

$$\eta(mt) = m^\alpha \eta(t); \qquad 0 < m < 1; \quad 0 < \alpha \leq 1.$$

This is denoted by SML(α, λ). Then the semi–Mittag-Leffler tailed distribution denoted by SMLT$(\alpha, \lambda, \theta)$ has Laplace transform

$$\phi_x(t) = \frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + \eta(t)}.$$

The first order semi-Mittag-Leffler tailed autoregressive (SMLTAR(1)) process has innovations whose Laplace transform is given by

$$\phi_\epsilon(t) = \left[\frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + \eta(t)}\right]\left[\frac{\lambda^\alpha + a^\alpha\eta(t)}{\lambda^\alpha + c\eta(t)}\right]$$

where $c = a^\alpha[p + (1-p)\theta]$. Therefore

$$\begin{aligned}\phi_\epsilon(t) &= \left[\frac{\lambda^\alpha + a^\alpha\eta(t)}{\lambda^\alpha + \eta(t)}\right]\left[\frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + c\eta(t)}\right] \\ &= \left[a^\alpha + (1-a^\alpha)\frac{\lambda^\alpha}{\lambda^\alpha + \eta(t)}\right]\left[\frac{\theta}{c} + \left(1 - \frac{\theta}{c}\right)\frac{\lambda^\alpha/c}{\lambda^\alpha/c + \eta(t)}\right].\end{aligned}$$

Therefore, the innovations $\{\epsilon_n\}$ can be represented as the sum of two independent semi–Mittag-Leffler tailed random variables u_n and v_n where

$$u_n \stackrel{d}{=} SMLT(\alpha, \lambda, a^\alpha) \text{ and } v_n \stackrel{d}{=} SMLT(\alpha, \lambda', \theta') \tag{2.5.2}$$

where $\lambda' = \lambda/c^{1/\alpha}, \theta' = \theta/c$. Then we have the following theorem which gives the stationary solution of the SMLTAR(1) model.

Theorem 2.5.1. *For* $0 < p < 1$, $0 < a < 1$ *the stationary Markov process* $\{x_n\}$ *defined by* (2.4.7) *has a semi–Mittag-Leffler tailed SMLT*$(\alpha, \lambda, \theta)$ *marginal distribution if and only if the innovation sequence* $\{\epsilon_n\}$ *are independent and identically distributed as the sum of two independent semi–Mittag-Leffler Tailed random variables as in* (2.5.2)*, provided* $x_0 \stackrel{d}{=} SMLT(\alpha, \lambda, \theta)$.

The stationarity of the process can be easily established, as given below.

$$\begin{aligned}
\phi_{x_n}(t) &= \phi_{\epsilon_n}(t)[p + (1-p)\phi_{x_{n-1}}(at)] \\
&= \left[\frac{\lambda^\alpha + a^\alpha \eta(t)}{\lambda^\alpha + \eta(t)}\right]\left[\frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + c\eta(t)}\right] \\
&\times \left[p + (1-p)\frac{\lambda^\alpha + \theta\eta(at)}{\lambda^\alpha + \eta(at)}\right] \\
&= \left[\frac{\lambda^\alpha + a^\alpha \eta(t)}{\lambda^\alpha + \eta(t)}\right]\left[\frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + c\eta(t)}\right]\left[\frac{\lambda^\alpha + c\eta(t)}{\lambda^\alpha + \eta(at)}\right] \\
&= \frac{\lambda^\alpha + \theta\eta(t)}{\lambda^\alpha + \eta(t)} \quad \text{since } \eta(at) = a^\alpha \eta(t).
\end{aligned}$$

Hence x_n is distributed as SMLT$(\alpha, \lambda, \theta)$. The necessity part follows easily from the derivation of the structure of the innovation sequence. Now we consider the following theorem.

Theorem 2.5.2. *In a positive valued stationary Markov process* $\{x_n\}$ *satisfying the first order autoregressive equation* $x_n = ax_{n-1} + \epsilon_n$, $0 < a < 1$ *the innovations* $\{\epsilon_n\}$ *are independently and identically distributed as a tailed distribution of the same type as that of* $\{x_n\}$ *if and only if* $\{x_n\}$ *are distributed as semi–Mittag-Leffler .*

Proof 2.5.1. We have, assuming stationarity,

$$\phi_x(t) = \phi_x(at)\phi_\epsilon(t).$$

Suppose

$$\phi_\epsilon(t) = \theta + (1-\theta)\phi_x(t) \text{ where } 0 \le \theta < 1.$$

Then

$$\phi_x(t) = \phi_x(at)[\theta + (1-\theta)\phi_x(t)].$$

Writing

$$\phi_x(t) = \frac{1}{1+\eta(t)}, \qquad \text{we get}$$

$$\begin{aligned}\frac{1}{1+\eta(t)} &= \frac{1}{1+\eta(at)}\left[\theta + (1-\theta)\frac{1}{1+\eta(t)}\right] \\ &= \left[\frac{1}{1+\eta(at)}\right]\left[\frac{1+\theta\eta(t)}{1+\eta(t)}\right].\end{aligned}$$

This implies $\eta(at) = \theta\eta(t)$. By taking $\theta = a^\alpha$, this means that the distribution of x_n is semi–Mittag-Leffler .

Conversely, if $\{x_n\}$ are semi–Mittag-Leffler , we get

$$\begin{aligned}\phi_\epsilon(t) &= \frac{\phi_x(t)}{\phi_x(at)} = \frac{1+\eta(at)}{1+\eta(t)} \\ &= \frac{1+a^\alpha\eta(t)}{1+\eta(t)} \\ &= a^\alpha + (1-a^\alpha)\frac{1}{1+\eta(t)}.\end{aligned}$$

Hence $\{\epsilon_n\}$ is distributed as SMLT(α, a^α).

The SMLTAR(1) process can be regarded as generalizations of the EAR(1), NEAR(1), MLAR(1), NMLAR(1), TEAR(1), ETAR(1) and MLTAR(1) processes. These processes are useful to model non-negative time series data which exhibit zeros, as in the case of stream flow data of rivers that are dry during part of the year. They are useful for modelling life times of devices which have some probability for damage immediately when it is put to use.In a similar manner,the

models can be extended to the semi-α-Laplace case and its special cases. Also geometric Mittag-Leffler and geometric alpha-Lapace distributions and time series models can be developed.

Exercises 2.5.

2.5.1. Derive the Laplace transform of the exponential tailed distribution

2.5.2. Derive the innovation structure of the Mittag-Leffler tailed autoregressive process.

2.5.3. Examine whether the Mittag-Leffler tailed distribution is self decomposable.

2.5.4. Give a real life example where the exponential tailed distribution can be used for modelling.

2.5.5. Show that Laplace distribution belongs to the semi-α-Laplace family.

2.5.6. Define a geometric exponential distribution similar to the geometric stable distribution.

2.5.7. Try to develop a generalized Laplacian model, with characteristic function $\varphi_x(t) = \left(\frac{1}{1-\beta^2 t^2}\right)^{\alpha}$.

2.5.8. Develop the concept in geometric infinite divisibility by replacing addition by minimum in the case of g.i.d.

2.5.9. Develop an autoregressive minification structure by replacing addition by minimum in the standard AR(1) equation.

2.6. Marshall-Olkin Weibull Time Series Models

2.6.1. Introduction

The need for developing time series models having non - Gaussian marginal distributions has been long felt from the fact that many naturally occurring time series are non - Gaussian with Markovian structure. In recent years Tavares (1980), Yeh et al. (1988), Arnold and Robertson (1989), Pillai (1991), Alice and Jose (2004, 2005) and others have developed various autoregressive models with minification structure. The Weibull distribution, including exponential distribution play a central role in the modeling of survival or lifetime data and time series data of non-negative random variables such as hydrological data and wind velocity magnitudes. Lewis and McKenzie (1991), Brown et al (1984) note that although studies have shown that Weibull marginal distributions have been found adequate for wind velocity magnitudes, unfortunately 'no time series models have been rigorously developed for random variables possessing a Weibull distribution'. Wind power data are even more likely to need very long tailed marginal distributions. Again in reliability studies, sequences of times between failures are correlated and models with non-constant marginal hazard rate are needed to model them adequately.

2.6.2. Marshall-Olkin semi-Weibull distribution and its properties

We say that a random variable X with positive support has a semi-Weibull distribution and write $X \stackrel{d}{=} SW(\beta, \rho)$ if its survival function is given by

$$\bar{F}_x(x) = P(X > x) = exp(-\Psi(x)) \tag{2.6.1}$$

where $\Psi(x)$ satisfies the functional equation,

$$\rho\Psi(x) = \Psi(\rho^{\frac{1}{\beta}}x); \beta > 0, 0 < \rho < 1 \tag{2.6.2}$$

Equation (2.6.2) will give on iteration

$$\rho^n\Psi(x) = \Psi(\rho^{\frac{n}{\beta}}x).$$

On solving (2.6.2) we obtain $\Psi(x) = x^\beta h(x)$, where h(x) is periodic in ln x with period $\left(\frac{-2\pi\beta}{ln\ \rho}\right)$. For details see Jose (1994, 2005) .

We consider a new family of distributions introduced by Marshall and Olkin (1997). Considering a survival function $\bar{F}$, we get the one-parameter family of survival functions

$$\bar{G}(x;\alpha) = \frac{\alpha\bar{F}(x)}{[1-(1-\alpha)\bar{F}(x)]}; -\infty < x < \infty, 0 < \alpha < \infty. \tag{2.6.3}$$

It can be easily seen that when $\alpha = 1$, $\bar{G} = \bar{F}$.

Whenever F has a density, the family of survival functions given by $\bar{G}(x;\alpha)$ in (2.6.3) has easily computed densities. In particular, if F has a density f and hazard rate r_F, then G has the density g given by

$$g(x;\alpha) = \frac{\alpha f(x)}{\{1-(1-\alpha)\bar{F}(x)\}^2} \tag{2.6.4}$$

and hazard rate

$$r(x;\alpha) = \frac{r_F(x)}{(1-(1-\alpha)\bar{F}(x)}; -\infty < x < \infty \tag{2.6.5}$$

Substituting (2.6.1) in (2.6.3) we get a new family of distributions, which we shall refer to as the survival function of Marshall-Olkin semi-Weibull [MOSW (α,β,ρ)] family, whose survival function is given by

$$\bar{G}(x;\alpha) = \frac{\alpha}{e^{\Psi(x)}-(1-\alpha)}; x > 0, \alpha > 0.$$

The probability density function corresponding to G is given by

$$g(x;\alpha) = \frac{\alpha e^{\Psi(x)}\Psi'(x)}{[e^{\Psi(x)}-(1-\alpha)]^2}; x > 0, \alpha > 0.$$

The hazard rate is given by

$$r(x;\alpha) = \frac{\Psi'(x)}{1-(1-\alpha)e^{-\Psi(x)}}; x > 0, \alpha > 0.$$

Now we establish the following properties.

Theorem 2.6.1. *Let N be an integer valued random variable independent of the X_n's such that $P[N \geq 2] = 1$ where $\{X_n\}$ is a sequence of independent and identically distributed MOSW random variables. Then $Y = (\frac{N}{\alpha})^{\frac{1}{\beta}}$ min $(X_1, X_2, ..., X_N)$; $N > \alpha, N > 1$ is distributed as semi-Weibull.*

Proof 2.6.1. We have

$$\begin{aligned}\bar{F}_Y(x) &= P[Y > x] \\ &= \sum_{n=2}^{\infty} P[N = n].P[Y > x|N = n] \\ &= \sum_{n=2}^{\infty} P[N = n].[\bar{F}_X((\frac{n}{\alpha})^{-1/\beta}x)]^n \\ &= \sum_{n=2}^{\infty} P[N = n].\left[\frac{1}{1 + (\frac{1}{\alpha})\Psi((\frac{n}{\alpha})^{-1/\beta}x)}\right]^n = e^{-\Psi(x)}\end{aligned}$$

Hence Y is distributed as semi-Weibull.

Theorem 2.6.2. *If $\{X_1, X_2, \cdots, X_n\}$ are independently and identically distributed as MOSP (α, β, p), then $Z_n = (\frac{n}{\alpha})^{\frac{1}{\beta}}$ $min(X_1, X_2, \cdots, X_n); \alpha, \beta > 0, n > 1, n > \alpha$;is asymptotically distributed as semi-Weibull.*

Proof 2.6.2. If X is distributed as Marshall-Olkin semi-Pareto, MOSP (α, β, p), then

$$\bar{F}(x; \alpha, \beta, p) = \frac{1}{1 + \frac{1}{\alpha}\psi(x)}$$

where

$$\psi(x) = \psi(p^{\frac{1}{\beta}}x).$$

Hence

$$\begin{aligned}\bar{F}_{z_n}(x) &= P\left[(\frac{n}{\alpha})^{\frac{1}{\beta}}(\min X_1, X_2, \cdots, X_n) > x\right] \\ &= \left[\bar{F}_X\left((\frac{n}{\alpha})^{-\frac{1}{\beta}}x\right)\right]^n \\ &= \left[\frac{1}{1 + \frac{\psi(x)}{n}}\right]^n \to e^{-\psi(x)}\end{aligned}$$

as n tends to infinity.

Similar results can be obtained in the case of Marshall Olkin Pareto and Weibull distribution as a special case.

Theorem 2.6.3. *Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed random variables with common survival function $\bar{F}(x)$ and N be a geometric random variable with parameter p and $P(N = n) = pq^{n-1}$; n=1,2,..., $0 < p < 1$, q=1-p, which is independent of $\{X_i\}$ for all $i \geq 1$. Let $U_N = \min_{1 \leq i \leq N} X_i$. Then $\{U_N\}$ is distributed as MOSW if and only if $\{X_i\}$ is distributed as semi-Weibull.*

Proof 2.6.3.

$$\begin{aligned} \bar{H}(x) &= P(U_N > x) \\ &= \sum_{n=1}^{\infty} [\bar{F}(x)]^n pq^{n-1} \\ &= \frac{p\bar{F}(x)}{1-(1-p)\bar{F}(x)}. \end{aligned}$$

Suppose

$$\bar{F}(x) = \exp(-\Psi(x)).$$

Then

$$\bar{H}(x) = \frac{1}{1 + (\frac{1}{p})(e^{\Psi(x)} - 1)},$$

which is the survival function of MOSW. This proves the sufficiency part of the theorem. Conversely, suppose

$$\bar{H}(x) = \frac{1}{1 + (\frac{1}{p})(e^{\Psi(x)} - 1)}.$$

Then we get

$$\bar{F}(x) = \exp(-\Psi(x)),$$

which is the survival function of semi-Weibull.

2.6.3. An AR(1) model with MOSW marginal distribution

In this section we consider a first order autoregressive model.

Theorem 2.6.4. *Consider an AR (1) structure given by*

$$X_n = \begin{cases} \epsilon_n & w.p \quad p \\ \min(X_{n-1}, \epsilon_n) & w.p \quad (1-p) \end{cases} \tag{2.6.6}$$

where $\{\epsilon_n\}$ is a sequence of independent and identically distributed random variables independent of X_n, then $\{X_n\}$ is a stationary Markovian AR(1) process with MOSW marginals if and only if $\{\epsilon_n\}$ is distributed as semi-Weibull distribution.

Proof 2.6.4. From (2.6.6) it follows that

$$\bar{F}_{X_n}(x) = p\bar{F}_{\epsilon_n}(x) + (1-p)\bar{F}_{X_{n-1}}(x).\bar{F}_{\epsilon_n}(x) \tag{2.6.7}$$

Under stationary equilibrium,

$$\bar{F}_X(x) = \frac{p\bar{F}_\epsilon(x)}{[1-(1-p)\bar{F}_\epsilon(x)]}.$$

If we take $\bar{F}_\epsilon(x) = e^{-\Psi(x)}$, then it easily follows that

$$\bar{F}_X(x) = \frac{p}{e^{\Psi(x)} - (1-p)},$$

which is the survival function of MOSW.
Conversely, if we take,

$$\bar{F}_{X_n}(x) = \frac{p}{e^{\Psi(x)} - (1-p)},$$

it is easy to show that $F_{\epsilon_n}(x)$ is distributed as semi-Weibull and the process is stationary. In order to establish stationarity we proceed as follows.
Assume $X_{n-1} \stackrel{d}{=}$ MOSW and $\epsilon_n \stackrel{d}{=}$ semi-Weibull.
Then

$$\bar{F}_{X_n}(x) = \frac{pe^{-\Psi(x)}}{1-(1-p)e^{-\Psi(x)}}.$$

This establishes that $\{X_n\}$ is distributed as MOSW. Even if X_0 is arbitrary, it is easy to establish that $\{X_n\}$ is stationary and is asymptotically marginally distributed as MOSW.

The following theorem is regarding a k^{th} order autoregressive model.

Theorem 2.6.5. *Consider an autoregressive model of order k as follows*

$$X_n = \begin{cases} \in_n & w.p. \quad p_0 \\ \min(X_{n-1}, \in_n) & w.p. \quad p_1 \\ \min(X_{n-2}, \in_n) & w.p. \quad p_2 \\ \vdots & \\ \min(X_{n-k}, \in_n) & w.p. \quad p_k \end{cases} \tag{2.6.8}$$

where $0 < p_i < 1, (p_1 + p_2 + \cdots + p_k) = 1 - p_0$. *Then* $\{X_n\}$ *has stationary marginal distribution as MOSW if and only if* $\{\epsilon_n\}$ *is distributed as semi-Weibull.*

The proof follows from the following facts.

$$\bar{F}_{X_n}(x) = p_0\, \bar{F}_{\epsilon_n}(x) + p_1\, \bar{F}_{X_{n-1}}(x)\, \bar{F}_{\epsilon_n}(x) + \cdots + p_k\, \bar{F}_{X_{n-k}}(x)\bar{F}_{\epsilon_n}(x)$$

Under stationary equilibrium,

$$\bar{F}_X(x) = p_0\, \bar{F}_\epsilon(x) + p_1\, \bar{F}_X(x)\, \bar{F}_\epsilon(x) + \cdots + p_k\, \bar{F}_X(x)\, \bar{F}_\epsilon(x)$$

This reduces to

$$\bar{F}_X(x) = \frac{p_0\, \bar{F}_\epsilon(x)}{[1 - (1 - p_0)\, \bar{F}_\epsilon(x)]}.$$

It can be seen that the semi-Weibull distribution is a more general class of distributions which includes Weibull distribution in the sense that for $h(x) = 1$, we have $\Psi(x) = x^\beta$.

2.6.4. Marshall-Olkin generalized Weibull distribution

Consider the two-parameter Weibull distribution with survival function

$$\bar{F}(x) = \exp(-(\lambda x)^\beta);\ x > 0, \lambda > 0, \beta > 0.$$

Then substituting in (5.2.3) we get a new family of distributions, which we shall refer to as the Marshall-Olkin Generalized Weibull (MOGW) family, whose survival function is given by

$$\bar{G}(x; \alpha, \lambda, \beta) = \frac{\alpha \exp[-(\lambda x)^\beta]}{1 - (1 - \alpha)\exp[-(\lambda x)^\beta]},\ x > 0, \lambda, \beta, \alpha > 0.$$

The probability density function corresponding to G is given by

$$g(x; \alpha, \lambda, \beta) = \frac{\alpha\beta\lambda^\beta x^{\beta-1} \exp(\lambda x)^\beta}{\exp[(\lambda x)^\beta - (1 - \alpha)]^2},\ x > 0, p, \beta, \alpha > 0.$$

The hazard rate is given by

$$r(x; p, \alpha, \beta) = \frac{\lambda^\beta \beta (x)^{\beta-1} \exp(\lambda x)^\beta}{\{\exp(\lambda x)^\beta - (1 - \alpha)\}},\ x > 0, \lambda, \beta, \alpha > 0.$$

We also explore the nature of the hazard rate $r(x)$. It is increasing if $\alpha \geq 1$, $\beta \geq 1$ and decreasing if $\alpha < 1$, $\beta < 1$. If $\beta > 1$, then r(x) is initially increasing and eventually increasing, but there may be an interval where it is decreasing.

Similarly if $\beta < 1$, then r(x) is initially decreasing and eventually increasing but there is an interval where it is increasing. When $\alpha = 1$, it coincides with the Weibull distribution. This points out the wide applicability of the MOGW distribution for modeling various types of reliability data. Theorem 2.6.4 and theorem 2.6.5 can be extended in this case also.

2.6.5. An AR (1) Model with MOGW marginal distribution

Theorem 2.6.6. *Consider the AR (1) structure given by*

$$X_n = \begin{cases} \epsilon_n & w.p. \quad p \\ \min(X_{n-1}, \epsilon_n) & w.p. \quad (1-p) \end{cases} \tag{2.6.9}$$

where $\{\in_n\}$ is a sequence of independent and identically distributed random variables independent of X_n ; then $\{X_n\}$ is a stationary Markovian AR (1) process with MOGW (p, λ, β) marginals if and only if $\{\in_n\}$ is distributed as Weibull distribution with parameters λ and β.

Proof 2.6.5. Proceeding as in the case of theorem 2.6.4 if we take

$$\bar{F}_{\epsilon}(x) = \exp(-\lambda x)^{\beta},$$

then it easily follows that

$$\begin{aligned} \bar{F}_X(x) &= \frac{p \exp(-\lambda x)^{\beta}}{[1-(1-p)exp(-\lambda x)^{\beta}]} \\ &= \frac{p}{[\exp(\lambda x)^{\beta} - (1-p)]} \end{aligned}$$

which is the survival function of MOGW (p, λ, β).
Conversely, if we take,

$$\bar{F}_{X_n}(x) = \frac{p}{[\exp(\lambda x)^{\beta} - (1-p)]},$$

it is easy to show that $\bar{F}_{\epsilon_n}(x)$ is distributed as Weibull with parameters λ, β and the process is stationary. In order to establish stationarity we proceed as follows. Assume $X_{n-1} \stackrel{d}{=} MOGW(p, \lambda, \beta)$ and $\{\in_n\} \stackrel{d}{=}$Weibull (λ, β).
Then

$$\bar{F}_{X_n}(x) = \frac{p[\exp(-\lambda x)^{\beta}]}{\{1-(1-p)\exp(-\lambda x)^{\beta}\}}.$$

This establishes that $\{X_n\}$ is distributed as $MOGW(p, \lambda, \beta)$. Even if X_0 is arbitrary, it is easy to establish that $\{X_n\}$ is stationary and is asymptotically marginally distributed as $MOGW(p, \lambda, \beta)$.

Theorem 2.6.7. *Consider an autoregressive model X_n of order k with structure (2.6.8). Then $\{X_n\}$ has stationary marginal distribution as MOGW if and only if $\{\epsilon_n\}$ is distributed as Weibull.*

Proof is similar to Theorem 2.6.9

Table 1 shows $P(X_n < X_{n-1})$, which are obtained through a Monte Carlo simulation procedure. Sequences of 100, 300, 500, 700, 900 observations from MOGWAR (1) process are generated repeatedly for ten times and for each sequence the probability is estimated. A table of such probabilities is provided with the average from ten trials along with an estimate of standard error in brackets. (see Table 1).

Table 1. $P(X_n < X_{n-1})$ for the MOGWAR(1) process where $\lambda = 1, \beta = 5$.

Sample size

p / n	200	400	600	800	1000
0.1	0.7705171 (0.002001358)	0.7523951 (0.003661458)	0.7573636 (0.2899318)	0.745249 (0.002795277)	0.7589128 (0.002238917)
0.2	0.6740114 (0.005064845)	0.6597776 (0.003381987)	0.6564568 (0.003264189)	0.6621319 (0.002337129)	0.6654536 (0.001765257)
0.3	0.6253248 (0.0551385)	0.5887756 (0.002164813)	0.5930888 (0.0259571)	0.5941391 (0.002407355)	0.5913183 (0.003051416)
0.4	0.5027496 (0.005699338)	0.4992585 (0.004394413)	0.5230972 (0.003465077)	0.5237219 (0.001724348)	0.5153204 (0.003295175)
0.5	0.4097892 (0.003480096)	0.4390949 (0.004124351)	0.4486389 (0.003437423)	0.4265652 (0.003071082)	0.4378709 (0.001833982)
0.6	0.3228981 (0.005857569)	0.3585286 (0.005020905)	0.3523585 (0.003365158)	0.3440362 (0.002421289)	0.357975 (0.003009052)
0.7	0.2695292 (0.005099691)	0.2676458 (0.003165013)	0.2646377 (0.003349412)	0.2773244 (0.003459127)	0.2731059 (0.003940234)
0.8	0.1873956 (0.005808668)	0.207557 (0.006569641)	0.1761628 (0.002695993)	0.1863542 (0.0033855295)	0.2003661 (0.003308759)
0.9	0.1175264 (0.007946155)	0.1194025 (0.006232362)	0.1119517 (0.005194928)	0.1012818 (0.002785095)	0.1130184 (0.003507806)

2.6.6. Case study

In this section, we illustrate the application of the MOGWAR(1) process in modeling a hydrology data as a case study. The data consists of total daily weighted discharge (in mm^3) of Neyyar river in Kerala at the location Amaravilla (near Amaravilla bridge) during 1993. Neyyar is one of the west flowing rivers in Kerala, located in the Southern most part. It originates from Agasthyamala at an elevation of about 1,860 m. above mean sea level. From there it flows down rapidly along steep slopes in its higher reaches and then winds its way through flat country in the lower reaches. In the initial stages the course is in a southwestern direction but at Ottasekharamangalam the river turns and flows west. It again takes a southwestern course from Valappallikanam upto its fall. The Neyyar is 56 Km. long and has a total drainage area of 497 sq. Km. It is a main source of irrigation in southern Kerala and the Neyyar Dam is a main source of hydroelectric power.

The arithmetic mean of the given data is 0.81. The estimates are obtained as $p = .5$ and $\beta = .7$. The calculated value of χ^2 is 0.626, which is significantly less than the tabled value. Hence MOGW distribution is found to be a good fit in this situation. It is found that the simulated MOGWAR (1) process has close resemblance to the actual data.

Exercises 2.6.

2.6.1. Define a minification process of order 1.

2.6.2. Obtain the class of distributions for which a stationary minification process is defined.

2.6.3. Develop a minification process with Pareto marginals.

2.6.4. Develop a semi-Weibull minification process.

2.6.5. Obtain the relationship between semi-Weibull and semi-Pareto distributions.

2.6.6. Obtain the innovation structure of a general Marshall-Olkin minification process.

2.6.7. Develop a bivariate Pareto minification process.

2.6.8. Develop a bivariate exponential minification process of order.

2.6.9. Derive the hazard rate function of a Marshall-Olkin exponential distribution.

2.6.10. Derive the stationary solution of a k^{th} order minification process.

References

Alice, T. & Jose, K.K. (2004). Bivariate semi-Pareto minification processes. *Metrika*, **59**, 305-313.

Alice, T. & Jose, K.K. (2005). Marshall-Olkin Semi-Weibull Minification Processes.*Recent Advances in Statistical Theory and Applications*, **1**, 6-17.

Anderson, D. N. and Arnold, B. C. (1993). Linnik distributions and processes, *J. Appl. Prob.*, **30**, 330-340.

Arnold, B. C. & Robertson, C. A. (1989). Autoregressive logistic processes. *J. Appl. Prob.*, **26**,524-531.

Berg, C. and Forst, G. (1975). *Potential Theory on Locally Compact Abelian Groups*, Springer, Berlin.

Bondesson, L. (1992). *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, Lecture Notes in Statistics **76**, Springer–Verlag, New York.

Brown, B. G.; Katz, R. W. & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *J. Climate Appl. Meteorol.* **23**, 1184-1195.

Chernick, M. R., Daley, D. J. and Littlejohn, R. P. (1988). A time–reversibility relationship between two Markov chains with stationary exponential distributions, *J. Appl. Prob.*, **25**, 418–422.

Devroye, L. (1990). A note on Linnik's distribution, *Statist. Prob. Letters*. **9**, 305–306.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York.

Fujita, Y. (1993). A generalization of the results of Pillai, *Ann. Inst. Statist. Math.*, **45**, 2, 361–365.

Gaver, D. P. and Lewis, P. A. W. (1980). First order autoregressive gamma sequences and point processes, *Adv. Appl. Prob.*, **12**, 727–745.

Jayakumar, K. and Pillai, R. N. (1993). The first order autoregressive Mittag-Leffler processes, *J. Appl. Prob.*, **30**, 462–466.

Jose, K. K. (1994). Some aspects of non-Gaussian autoregressive time series modeling, Unpublished Ph.D. Thesis submitted to University of Kerala.

Jose, K. K. (2005). Autoregressive models for time series data with exact zeros, submitted.

Karlin, S.and Taylor, E. (2002) *A first course in stochastic processes*, Academic press, London.

Klebanov, L. B., Maniya, G. M. and Melamed, I. A. (1984). A problem of Zolotarev and analogues of infinitely divisible and stable distributions in a scheme for summing a random number of random variables, *Theory Probab. Appl.*, **29**, 791–794.

Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory*, John Wiley and Sons, New York.

Lawrance, A. J. (1991). Dierectionality and reversibility in time series. *Int.Stat.Review*, **59(1)**, 67-79. Lawrance, A. J. and Lewis, P. A. W. (1981). A new autoregressive

time series model in exponential variables (NEAR(1)), *Adv. Appl. Prob.*, **13**, 826–845.

Lewis, P. A. W. and McKenzie, E. (1991). Minification processes and their transformations. *J. Appl. Prob.*, **28**, 45-57.

Littlejohn, R. P. (1993). A reversibility relation for two Markovian time series models with stationary exponential tailed distribution, *J. Appl. Prob.*, (Preprint).

Marshall, A. W. & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrica,* **84**, 3, 641-652.

Medhi, A. (2006) *Stochastic models in queueing theory,* Academic Press,California.

Medhi, A. (2004) *Stochastic processes*, New Age Publishers, New Delhi.

Mohan, N. R., Vasudeva, R. and Hebbar, H. V. (1993). On geometrically infinitely divisible laws and geometric domains of attraction, *Sankhya.* **55** A, 2, 171–179.

Papoulis, E. (2000). *Probability, random variables and stochastic processes*, McGraw Hill Publishers, New York.

Pillai, R. N. (1971). Semi stable laws as limit distributions, *Ann. Math. Statist.*, **42**, 2, 780–783.

Pillai, R. N. (1985). Semi–α–Laplace distributions, *Commun. Statist.–Theor. Meth.*, **14**(4), 991–1000.

Pillai, R. N. (1990a). On Mittag-Leffler functions and related distributions, *Ann. Inst. Statist. Math.*, **42**, 1, 157–161.

Pillai, R. N. (1990b). Harmonic mixtures and geometric infinite divisibility, *J. Indian Statist. Assoc.*, **28**, 87–98.

Pillai, R. N. (1991). Semi–Pareto processes, *J. Appl. Prob.*, **28**, 461–465.

Pillai, R. N. and Sandhya, E. (1990). Distributions with complete monotone derivative and geometric infinite divisibility, *Adv. Appl. Prob.*, **22**, 751–754.

Rachev, S. T. and SenGupta, A. (1992). Geometric stable distributions and Laplace–Weibull mixtures, *Statistics and Decisions*, **10**, 251–271.

Ross, S. M. (2002). *Probability models in stochastics*, Academic Press, New Delhi.

Steutel, F.W. (1979). Infinite divisibility in theory and practice, *Scand. J. Statist.*, **6**, 57–64.

Steutel, F. W. and Van Harn, K. (1979). Discrete analogues of self–decomposability and stability, *Ann. Prob.,* **7**, 893–899.

Tavares, L. V. (1980). An exponential Markovian stationary process. *J. Appl. Prob.*, **17**, 1117-1120.

Yeh, H. C.; Arnold, B. C. & Robertson, C. A. (1988). Pareto processes. *J. Appl. Prob.,* **25**, 291-301.

2.7. On Concomitants of Order Statistics and Concomitants of Record Values: Applications in Point Estimation

[*This Section is based on the lectures of Professor P. Yageen Thomas of the Department of Statistics, University of Kerala.*]

2.7.0. Introduction

Order statistics deal with properties and applications of ordered random variables and their functions. Order statistics play a very important role in statistical theory as it helps to develop methods of statistical inference which are valid with respect to a broad class of population distribution functions. In several situations, methods based on order statistics are proved to be most efficient when compared with others. These methods are widely accepted due to their simplicity and robustness, even at the cost of some loss of efficiency.

Since there is no direct extension of order concept to multivariate random variables, the extension of procedure based on order statistics to such situations is inapplicable. But however from a random sample arising from a bivariate distribution, ordering of the values recorded on the first variable generates a set of random variables associated with the corresponding Y variate. These random variables obtained due to the ordering of the $X's$ are known as the concomitants of order statistics. Let (X, Y) be a random vector with joint cumulative distribution function $(cdf)F(x, y)$ and joint probability density function $(pdf)f(x, y)$. Let (X_i, Y_i), $i = 1, 2, ..., n$ be a random sample drawn from the distribution of (X, Y). Let $X_{i:n}$ be the i^{th} order statistic of the X observation, then the Y variate

associated with the $X_{i:n}$ is called the concomitant of the i^{th} order statistic and is denoted by $Y_{[i:n]}$. It may be noted that Bhattacharya (1974) has independently developed the above concept of concomitants of order statistics and he called them as induced order statistics.

Applications of concomitants of order statistics arises in several problems of study. The most important use of concomitants of order statistics arises in selection procedures when $k(< n)$ individuals are chosen on the basis of their X-values. Then the corresponding Y-values represent the performance on an associated characteristic. For example, if the top k out of n bulls, as judged by their genetic make up, are selected for breeding, then $Y_{[n-k+1:n]}, \cdots, Y_{[n:n]}$ might represent the average milk yield of their female offspring. As another example, X might be the score on a screening test and Y the score on a latter test. In this example only the top k performers in the screening test are selected for further training and their scores on a second test generates the concomitants of order statistics. These concomitants of order statistics help one to reduce the complexity of identifying the best performers among a group of individuals.

Suppose the parent bivariate distribution is defined with $cdf\ \ F(x, y)$ and pdf $f(x, y)$, then the pdf of the r^{th} concomitant $Y_{[r:n]}$ for $1 \leq r \leq n$ is given by (see, David and Nagaraja, 2003, p.144),

$$g_{[r:n]}(y) = \int_x f(y|x) f_{r:n}(x) \mathrm{d}x, \tag{2.7.1}$$

where $f_{r:n}(x)$ is the pdf of the r^{th} order statistic $X_{r:n}$ of the X variate and $f(y|x)$ is the conditional pdf of Y given $X = x$.
The joint pdf of $Y_{[r:n]}$ and $Y_{[s:n]}$ for $1 \leq r < s \leq n$ is given by (see, David and Nagaraja, 2003, p.144),

$$g_{[r,s:n]}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} f(y_1|x_1) f(y_2|x_2) f_{r,s:n}(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2, \tag{2.7.2}$$

where $f_{r,s:n}(x_1, x_2)$ is the joint pdf of $X_{r:n}$ and $X_{s:n}$. From Yang (1977) we get the expressions for, $E(Y_{[r:n]})$, $\mathrm{Var}(Y_{[r:n]})$, for $1 \leq r \leq n$ and $\mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]})$ for $1 \leq r < s \leq n$ and are given below.

$$E(Y_{[r:n]}) = E[E(Y|X_{r:n})], \tag{2.7.3}$$

$$\mathrm{Var}(Y_{[r:n]}) = \mathrm{Var}[E(Y|X_{r:n})] + E[\mathrm{Var}(Y|X_{r:n})] \tag{2.7.4}$$

and

$$\mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \mathrm{Cov}[E(Y|X_{r:n}), E(Y|X_{s:n})]. \tag{2.7.5}$$

There is extensive literature available on the application of concomitants of order statistics such as in: biological selection problem (see, Yeo and David,1984), ocean engineering (see, Castillo,1988), development of structural designs (see, Coles and Tawn,1994) and so on. Concomitants of order statistics have been used by several authors in estimating the parameters of bivariate distributions. Harrell and Sen (1979) and Gill et al. (1990) have used concomitants of order statistics to estimate the parameters of a bivariate normal distribution. Spruill and Gastwirth (1982) have considered another interesting use of concomitants in estimating the correlation coefficient between two random variables X and Y. Barnett et al. (1976) have considered different estimators for the correlation coefficient of a bivariate normal distribution based on concomitants of order statistics. The distribution theory of concomitants in the bivariate Weibull distribution of Marshall and Olkin is discussed in Begum and Khan (2000a). Begum and Khan (2000b) have also developed the distribution theory of concomitants of order statistics from Gumbel's bivariate logistic distribution. In section 2, we consider an application of concomitants of order statistics in estimating a parameter of Morgenstern type bivariate uniform distribution.

Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be a sequence of independent and identically distributed random variables with cdf $F(x, y), (x, y) \in R \times R$. Let $F_X(x)$ and $F_Y(y)$ be the marginal $cdfs$ of X and Y respectively. Let $\{R_n, n \geq 1\}$ be the sequence of upper record values (see, Arnold et al., 1998, p.8) in the sequence of $X's$ as defined by,

$$R_n = X_{T_n}, \quad n = 1, 2, \cdots$$

where $T_1 = 1$ and $T_n = \min\{j : X_j > X_{T_{n-1}}\}$ for $n \geq 2$. Then the Y-variate associated with the X-value, which qualified as the n^{th} record will be called the concomitant of the n^{th} record and will be denoted by $R_{[n]}$. Suppose in an experiment, individuals are measured based on an inexpensive test and only those individuals whose measurement breaks the previous records are retained for the measurement based on an expensive test; then the resulting data involves concomitants of record values. For a detailed discussion on the distribution theory

of concomitants of record values see, Arnold et al. (1998) and Ahsanullah and Nevzorov (2000).

The pdf of n^{th} $(n \geq 1)$ record value is given by,

$$g_{R_n}(x) = \frac{1}{(n-1)!}[-log(1-F_X(x))]^{n-1} f_X(x) \tag{2.7.6}$$

and the joint pdf of m^{th} and n^{th} record values for $m < n$ is given by,

$$g_{R_m,R_n}(x_1,x_2) = \frac{[-log(1-F_X(x_1))]^{m-1}}{(m-1)!}\frac{[-log(1-F_X(x_2)) + log(1-F_X(x_1))]^{n-m-1}}{(n-m-1)!} \times \frac{f_X(x_1)f_X(x_2)}{1-F_X(x_1)}. \tag{2.7.7}$$

Thus the pdf of the concomitant of n^{th} record value is given by

$$f_{R_{[n]}}(y) = \int_{-\infty}^{\infty} f(y|x)g_{R_n}(x)\mathrm{d}x,$$

where $g_{R_n}(x)$ is as defined in (2.7.6) and $f(y|x)$ is the conditional pdf of Y given $X = x$ of the parent bivariate distribution.

The joint pdf of concomitants of m^{th} and n^{th} record values is given by (see, Ahsanullah and Nevzorov, 2000),

$$g_{R_{[m]},R_{[n]}}(y_1,y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{x_2} f(y_1|x_1)f(y_2|x_2)g_{R_m,R_n}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2,$$

where $g_{R_m,R_n}(x_1,x_2)$ is defined by (2.7.7). Some properties of concomitants of record values were discussed in Houchens (1984), Ahsanullah and Nevzorov (2000) and Arnold et al. (1998). However, not much work is seen done in the distribution theory and applications of concomitants of records in statistical inference problems. In subsection 2.7.2, we provide an application of concomitants of record values in estimating some parameters of Morgenstern type bivariate logistic distribution.

2.7.1. Application of concomitants of order statistics in estimating a parameter of Morgenstern type bivariate uniform distribution

Scaria and Nair (1999) have discussed the distribution theory of concomitants of order statistics arising from Morgenstern family of distributions (MFD) with cdf defined by (see, Kotz et al., 2000, P.52),

$$F(x,y) = F_X(x)F_Y(y)\{1+\alpha(1-F_X(x))(1-F_Y(y))\}, \quad -1 \le \alpha \le 1. \tag{2.7.8}$$

An important member of the MFD is Morgenstern type bivariate uniform distribution with pdf given by,

$$F(x,y) = \frac{xy}{\theta_1\theta_2}\left\{1+\alpha\left(1-\frac{x}{\theta_1}\right)\left(1-\frac{y}{\theta_2}\right)\right\}, \quad 0<x<\theta_1, \quad 0<y<\theta_2; \quad -1 \le \alpha \le 1. \tag{2.7.9}$$

Now we derive the Best Linear Unbiased Estimator (BLUE) of the parameter θ_2 involved in (2.7.9) using concomitants of order statistics (see, Chacko and Thomas, 2004).

Let $Y_{[r:n]}, \ r = 1, 2, \cdots, n$ be the concomitants of order statistics of a random sample of size n drawn from (2.7.9). Then the pdf of $Y_{[r:n]}$ and the joint pdf of $Y_{[r:n]}$ and $Y_{[s:n]}$ are obtained as,

$$g_{[r:n]}(y) = \frac{1}{\theta_2}\left[1+\alpha\frac{n-2r+1}{n+1}\left(1-\frac{2y}{\theta_2}\right)\right], \quad 1 \le r \le n. \tag{2.7.10}$$

and

$$\begin{aligned} g_{[r,s:n]}(y_1,y_2) = \frac{1}{\theta_2^2}\Big[1 &+ \alpha\frac{n-2r+1}{n+1}\left(1-\frac{2y_1}{\theta_2}\right) \\ &+ \alpha\frac{n-2s+1}{n+1}\left(1-\frac{2y_2}{\theta_2}\right) + \alpha^2\left(\frac{n-2s+1}{n+1} - \frac{2r(n-2s)}{(n+1)(n+2)}\right) \\ &\times\left(1-\frac{2y_1}{\theta_2}\right)\left(1-\frac{2y_2}{\theta_2}\right)\Big], \quad 1 \le r < s \le n. \end{aligned} \tag{2.7.11}$$

From (2.7.10) and (2.7.11) we get the means, variances and covariances of concomitants of order statistics as follows:

$$\begin{aligned} E[Y_{[r:n]}] &= \theta_2 \left[\frac{1}{2} - \alpha \frac{n-2r+1}{6(n+1)}\right] \\ &= \theta_2 \xi_{r:n}, \end{aligned} \tag{2.7.12}$$

where

$$\xi_{r:n} = \frac{1}{2} - \alpha \frac{n-2r+1}{6(n+1)}.$$

$$\begin{aligned} \text{Var}[Y_{[r:n]}] &= \theta_2^2 \left[\frac{1}{12} - \frac{\alpha^2 (n-2r+1)^2}{36(n+1)^2}\right] \\ &= \theta_2^2 \rho_{r,r:n}, \end{aligned} \tag{2.7.13}$$

where

$$\rho_{r,r:n} = \frac{1}{12} - \frac{\alpha^2 (n-2r+1)^2}{36(n+1)^2}$$

and

$$\begin{aligned} \text{Cov}[Y_{[r:n]}, Y_{[s:n]}] &= \theta_2^2 \frac{\alpha^2}{36} \left[\frac{(n-2s+1)}{(n+1)} - \frac{2r(n-2s)}{(n+2)(n+1)} - \frac{(n-2r+1)(n-2s+1)}{(n+1)^2}\right] \\ &= \theta_2^2 \rho_{r,s:n}, \end{aligned} \tag{2.7.14}$$

where

$$\rho_{r,s:n} = \frac{\alpha^2}{36} \left[\frac{(n-2s+1)}{(n+1)} - \frac{2r(n-2s)}{(n+2)(n+1)} - \frac{(n-2r+1)(n-2s+1)}{(n+1)^2}\right].$$

Let $\mathbf{Y}_{[n]} = [Y_{[1:n]}, \cdots, Y_{[n:n]}]'$ be the vector of concomitants. Then from (2.7.12) we can write

$$E(\mathbf{Y}_{[n]}) = \theta_2 \xi,$$

where

$$\xi = [\xi_{[1:n]}, \cdots, \xi_{[n:n]}]'.$$

Then from (2.7.13) and (2.7.14), the variance covariance matrix of $\mathbf{Y}_{[n]}$ is given by

$$D(\mathbf{Y}_{[n]}) = G\theta_2^2,$$

where

$$G = ((\rho_{r,s:n})).$$

If α is known then $(\mathbf{Y}_{[n]}, \theta_2\xi, \theta_2^2 G)$ is a generalized Gauss-Markov setup and hence the BLUE ($\hat{\theta_2}$ of θ_2) is given by,

$$\hat{\theta_2} = (\xi' G^{-1}\xi)^{-1}\xi' G^{-1}\mathbf{Y}_{[n]}$$

and the variance of $\hat{\theta_2}$ is given by,

$$\text{Var}(\hat{\theta_2}) = (\xi' G^{-1}\xi)^{-1}\theta_2^2.$$

It is clear that $\hat{\theta_2}$ is a linear function of the concomitants $Y_{[r:n]}$ $r = 1, 2, \cdots, n$. Hence we can write $\hat{\theta_2} = \sum_{r=1}^{n} a_r Y_{[r:n]}$, where a_r, $r = 1, 2, \cdots, n$ are constants. It is to be noted that the possible values of α are in the interval $[-1, 1]$. If the estimate $\hat{\theta_2}$ of θ_2 for a given $\alpha = \alpha_0 \in [-1, 1]$ is evaluated, then one need not consider the estimate for θ_2 for $\alpha = -\alpha_0$ as the coefficients of the estimate in this case can be obtained from the coefficients of $\hat{\theta_2}$ for $\alpha = \alpha_0$. This property can be easily observed from the following theorem:

Theorem 2.7.1. *Let $Y_{[r:n]}$, $r = 1, 2, \cdots, n$ be the concomitants of order statistics of a random sample $(X_i, Y_i), i = 1, 2, \cdots, n$ arising from (2.7.9) for a given $\alpha = \alpha_0 \in [-1, 1]$. Let the BLUE $\hat{\theta}_2(\alpha_0)$ of θ_2 for given α_0 based on the concomitants $Y_{[r:n]}, r = 1, 2, \cdots, n$ be written as $\hat{\theta}_2(\alpha_0) = \sum_{r=1}^{n} a_r Y_{[r:n]}$. Then the BLUE of $\hat{\theta}_2(-\alpha_0)$ of θ_2 when $\alpha = -\alpha_0$ is given by*

$$\hat{\theta}_2(-\alpha_0) = \sum_{r=1}^{n} a_{n-r+1} Y_{[r:n]} \quad \textit{with} \quad \text{Var}[\hat{\theta}_2(-\alpha_0)] = \text{Var}[\hat{\theta}_2(\alpha_0)].$$

Proof 2.7.1. From (2.7.13) and (2.7.14) for $1 \le r \le n$ we have $\rho_{r,r:n} = \rho_{n-r+1,n-r+1:n}$ and for $1 \le r < s \le n$, we have $\rho_{r,s:n} = \rho_{n-s+1,n-r+1:n}$. Moreover G is symmetric.

Therefore we can write for any $\alpha \in [-1, 1]$, $G = JGJ$, where J is an an $n \times n$ matrix given by,

$$J = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ . & . & \cdots & . \\ 1 & \cdots & 0 & 0 \end{bmatrix}$$

Again from (2.7.12) we have for any $\alpha_0 \in [-1, 1]$,

$$\begin{aligned} \xi_{r:n}(\alpha_0) &= \frac{1}{2} - \alpha_0 \frac{n - 2r + 1}{6(n+1)} \\ &= \xi_{n-r+1:n}(-\alpha_0). \end{aligned}$$

Thus

$$\xi(-\alpha_0) = J\xi(-\alpha_0).$$

Therefore, if $\alpha = \alpha_0$ is changed to $\alpha = -\alpha_0$ then the estimate $\hat{\theta}_2(-\alpha_0)$ is given by,

$$\begin{aligned} \hat{\theta}_2(-\alpha_0) &= (\xi'(-\alpha_0)G^{-1}\xi(-\alpha_0))^{-1}\xi'(-\alpha_0)G^{-1}Y_{[n]} \\ &= (\xi'(\alpha_0)JG^{-1}J\xi(\alpha_0))^{-1}\xi'(\alpha_0)JG^{-1}Y_{[n]}(-\alpha_0). \end{aligned}$$

Since $JJ = I$ and $JGJ = G$, we get,

$$\begin{aligned} \hat{\theta}_2(-\alpha_0) &= (\xi'(\alpha_0)G^{-1}\xi(\alpha_0))^{-1}\xi'(\alpha_0)G^{-1}JY_{[n]} \\ &= \sum_{r=1}^{n} a_r Y_{[n-r+1:n]}. \end{aligned}$$

That is the coefficient of $Y_{[r:n]}$ in $\hat{\theta}_2$ for $\alpha = \alpha_0$ is the same as the coefficient of $Y_{[n-r+1:n]}$ in $\hat{\theta}_2$ for $\alpha = -\alpha_0$. Similarly we get

$$\mathrm{Var}[\hat{\theta}_2(-\alpha_0)] = \mathrm{Var}[\hat{\theta}_2(\alpha_0)].$$

Thus the theorem is proved.

We have evaluated the coefficients a_r of $Y_{[r:n]}, 1 \le r \le n$ in $\hat{\theta}_2$ and $\mathrm{Var}(\hat{\theta}_2)$ for $n = 2(1)10$ and $\alpha = 0.25(0.25)0.75$ and are given in table 2.7.1. In order to

obtain the efficiency of our estimate $\hat{\theta}_2$, we introduce a simple unbiased estimate of θ_2 as,

$$\tilde{\theta}_2 = Y_{[1:n]} + Y_{[n:n]},$$

with variance given by,

$$\text{Var}(\tilde{\theta}_2) = \theta_2^2 \left[\frac{1}{6} + \frac{\alpha^2}{18} + \left(\frac{2n}{(n+1)(n+2)} - \frac{n-1}{n+1} \right) \right].$$

We have obtained the ratio $\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\tilde{\theta}_2)}$ as a measure of the efficiency $e_1 = e(\hat{\theta}_2|\tilde{\theta}_2)$ of our estimator $\hat{\theta}_2$ relative to the unbiased estimator $\tilde{\theta}_2$ for $n = 2(1)10$ and $\alpha = 0.25(0.25)0.75$. It can be seen that the efficiency of our estimator $\hat{\theta}_2$ of θ_2 is relatively very high when compared with $\tilde{\theta}_2$. An advantage of the above method of obtaining the BLUE of θ_2 is that with the expressions for $E[Y_{[r:n]}]$ and $\text{Cov}[Y_{[r:n]}, Y_{[s:n]}]$ one can also obtain without any difficulty the BLUE of θ_2 even if a censored sample alone is available.

2.7.2. Application of concomitants of record values in estimating some parameters of Morgenstern type bivariate logistic distribution

In this section we (see, Chacko and Thomas, 2005) consider the concomitants of record values arising from Morgenstern Family of Distributions with cdf given in (2.7.1). We further derive the joint pdf of concomitants of m^{th} and n^{th} $(m < n)$ record values arising from MFD. Based on these expressions we also derive the explicit expression for the product moments of concomitants of record values.

An important member of the MFD is the Morgenstern Type Bivariate Logistic Distribution ($MTBLD$) and its cdf is given by,

$$F_{X,Y}(x,y) = \left[1+\exp\left\{-\frac{x-\theta_1}{\sigma_1}\right\}\right]^{-1}\left[1+\exp\left\{-\frac{y-\theta_2}{\sigma_2}\right\}\right]^{-1}$$
$$\times\left[1+\alpha\left\{1-\left[1+\exp\left\{-\frac{x-\theta_1}{\sigma_1}\right\}\right]^{-1}\right\}\left\{1-\left[1+\exp\left\{-\frac{y-\theta_2}{\sigma_2}\right\}\right]^{-1}\right\}\right], \tag{2.7.15}$$
$$(x,y)\in R^2;\ \ (\theta_1,\theta_2)\in R^2;\ \ \sigma_1>0,\ \sigma_2>0,\ -1<\alpha<1$$

Suppose in certain complicated experiments significance is attributed to the values of the secondary measurement made by an accurate expensive test on individuals having record values with respect to the measurement made preliminarily on them by an inexpensive test. Now we derive (see, Chacko and Thomas 2005) the BLUE's of θ_2 and σ_2 involved in the $MTBLD$ defined by (2.7.15) when α is known and also obtained the BLUE of θ_2 when σ_2 and α are known based on concomitants of first n record values.

The joint cdf of the standard $MTBLD$ is obtained by making the transformation $u=\frac{x-\theta_1}{\sigma_1}$ and $v=\frac{y-\theta_2}{\sigma_2}$ in (2.7.15) and is given by,

$$F_{U,V}(u,v) = [1+\exp(-u)]^{-1}[1+\exp(-v)]^{-1}\left\{1+\alpha\frac{\exp(-u-v)}{[1+\exp(-u)][1+\exp(-v)]}\right\}. \tag{2.7.16}$$

Let $(U_i,V_i), i=1,2,\cdots$ be a sequence of independent observations drawn from (2.7.16). Let $R^*_{[n]}$ be the concomitant of the n^{th} record value$R^*_{[n]}$ arising from (2.7.16). Then the pdf $f^*_{[n]}(v)$ of $R^*_{[n]}$ and the joint pdf $f^*_{[m,n]}(v_1,v_2)$ of $R^*_{[m]}$ and $R^*_{[n]}$ for $m<n$ are given below,

$$f^*_{[n]}(v) = [1+\exp(-v)]^{-2}\exp(-v)\left\{1+\alpha(1-2^{1-n})\left[\frac{1-\exp(-v)}{1+\exp(-v)}\right]\right\}. \tag{2.7.17}$$

and for $m<n$,

$$
\begin{aligned}
f^{*}_{[m,n]}(v_1, v_2) = {} & [1+\exp(-v_1)]^{-2}\,[1+\exp(-v_2)]^{-2}\exp(-v_1-v_2) \\
& \times \Bigg[1 + \alpha\,\{2I_1(m,n)-1\}\left(\frac{1-\exp(-v_1)}{1+\exp(-v_1)}\right) \\
& + \alpha\,\{2I_2(m,n)-1\}\left(\frac{1-\exp(-v_2)}{1+\exp(-v_2)}\right) \\
& + \alpha^2\,\{4I_3(m,n) - 2I_1(m,n) - 2I_2(m,n) + 1\} \\
& \times \left(\frac{1-\exp(-v_1)}{1+\exp(-v_1)}\right)\left(\frac{1-\exp(-v_2)}{1+\exp(-v_2)}\right)\Bigg],
\end{aligned}
$$

where,

$$
\begin{aligned}
I_1(m,n) = {} & \frac{1}{(m-1)!(n-m-1)!}\sum_{r=0}^{n-m-1}(-1)^{n-m-r-1}\binom{n-m-1}{r} \\
& \times \left[\frac{(n-1)!}{n-r-1} - (n-r-2)!\,r! + (n-r-2)!\sum_{s=0}^{n-r-2}\frac{1}{s!}\frac{(r+s)!}{2^{r+s+1}}\right],
\end{aligned}
\tag{2.7.18}
$$

$$
I_2(m,n) = \frac{(n-1)!}{(m-1)!(n-m-1)!}\left(1-\frac{1}{2^n}\right)\sum_{r=0}^{n-m-1}(-1)^{n-m-r-1}\binom{n-m-1}{r}\frac{1}{n-r-1}
\tag{2.7.19}
$$

and

$$
\begin{aligned}
I_3(m,n) = {} & \frac{1}{(m-1)!(n-m-1)!}\sum_{r=0}^{n-m-1}(-1)^{n-m-r-1}\binom{n-m-1}{r}\Bigg[\frac{(n-1)!}{n-r-1}\left(1-\frac{1}{2^n}\right) \\
& - (n-r-2)\left(r!\left(1-\frac{1}{2^{r+1}}\right) - \sum_{s=0}^{n-r-2}\frac{(r+s)!}{s!}\left(\frac{1}{2^{r+s+1}} - \frac{1}{3^{r+s+1}}\right)\right)\Bigg].
\end{aligned}
\tag{2.7.20}
$$

Thus the means, variances and covariances of concomitants of first n record values (for $n \geq 1$) arising from (2.7.16) are given by,

$$E[R^*_{[n]}] = \alpha(1 - 2^{1-n}) = \mu_n \quad \text{(say)}, \tag{2.7.21}$$

$$\text{Var}[R^*_{[n]}] = \frac{\pi^2}{3} - \alpha^2(1 - 2^{1-n})^2 = \nu_{n,n} \quad \text{(say)} \tag{2.7.22}$$

and for $m < n$,

$$\begin{aligned}\text{Cov}[R^*_{[m]},\ R^*_{[n]}] &= \alpha^2[\{4I_3(m,n) - 2I_1(m,n) - 2I_2(m,n) + 1\} - (1 - 2^{1-m})(1 - 2^{1-n})] \\ &= V_{m,n} \quad \text{(say)}\ , \end{aligned} \tag{2.7.23}$$

Let $(X_i, Y_i)\quad i = 1, 2, \cdots$ be a sequence of independent observations drawn from a population with cdf defined by (2.7.15). If we write $u = \frac{x-\theta_1}{\sigma_1}$ and $v = \frac{y-\theta_2}{\sigma_2}$ then we have $X_i = \theta_1 + \sigma_1 U_i$ and $Y_i = \theta_2 + \sigma_2 V_i$ for $i = 1, 2, \cdots$. Then by using (2.7.21), (2.7.22) and (2.7.23) we have for $n \geq 1$,

$$E[R_{[n]}] = \theta_2 + \sigma_2\mu_n, \tag{2.7.24}$$

$$\text{Var}[R_{[n]}] = \sigma_2^2\nu_{n,n} \tag{2.7.25}$$

and for $m < n$,

$$\text{Cov}[R_{[m]},\ R_{[n]}] = \sigma_2^2\nu_{m,n}, \tag{2.7.26}$$

Clearly from (2.7.20), (2.7.21) and (2.7.22) it follows that $\mu_n, \nu_{n,n}$ and $\nu_{m,n}$ are known constants provided α is known. Suppose $\mathbf{R}_{[n]} = (R_{[1]}, R_{[2]}, \cdots, R_{[n]})$ denote the vector of concomitants of first n record values. Then from (2.7.24) to (3.7.26), we can write

$$E[\mathbf{R}_{[n]}] = \theta_2 1 + \sigma_2\mu, \tag{2.7.27}$$

where 1 is a column vector of n ones and $\mu = (\mu_1, \cdots, \mu_n)'$. Then the variance-covariance matrix of $\mathbf{R}_{[n]}$ is given by,

$$D[\mathbf{R}_{[n]}] = H\sigma_2^2, \tag{2.7.28}$$

where $H = ((\nu_{i,j}))$. If α involved in μ and H are known, then (2.7.27) and (2.7.28) together defines a generalized Gauss-Markov setup and then (proceeding as in David and Nagaraja 2003, p. 185) the BLUE's of θ_2 and σ_2 are given by

$$\hat{\theta}_2 = \frac{\mu' H^{-1}(\mu 1' - 1\mu')H^{-1}}{\Delta}\mathbf{R}_{[n]} \tag{2.7.29}$$

and

$$\hat{\sigma}_2 = \frac{1' H^{-1}(1\mu' - \mu 1')H^{-1}}{\Delta}\mathbf{R}_{[n]}, \tag{2.7.30}$$

where

$$\Delta = (\mu' H^{-1}\mu)(1' H^{-1}1) - (\mu' H^{-1}1)^2$$

The variances of the above estimators are given by

$$\text{Var}(\hat{\theta}_2) = \left(\frac{\mu' H^{-1}\mu}{\Delta}\right)\sigma_2^2, \tag{2.7.31}$$

and

$$\text{Var}(\hat{\sigma}_2) = \left(\frac{1' H^{-1}1}{\Delta}\right)\sigma_2^2. \tag{2.7.32}$$

Clearly $\hat{\theta}_2$ and $\hat{\sigma}_2$ can be written as $\hat{\theta}_2 = \sum_{i=1}^{n} b_i R_{[i]}$ and $\hat{\sigma}_2 = \sum_{i=1}^{n} c_i R_{[i]}$ where b_i and c_i, $i = 1, 2, \cdots, n$ are constants.

We have evaluated, the coefficients b_i and c_i of $R_{[i]}$, $1 \le i \le n$ in $\hat{\theta}_2$ and $\hat{\sigma}_2$; $\text{Var}(\hat{\theta}_2)$ and $\text{Var}(\hat{\sigma_2})$ for $n = 2(1)10$ and $\alpha = 0.25(0.25)0.75$ and are given in table 2.7.2 and table 2.7.3 respectively. In order to compare the efficiencies of our estimators $\hat{\theta}_2$ and $\hat{\sigma}_2$ we introduce two simple unbiased estimators of θ_2 and σ_2 based on the concomitants of the first and n^{th} records as given below,

$$\tilde{\theta}_2 = R_{[1]}$$

and

$$\tilde{\sigma}_2 = \frac{R_{[n]} - R_{[1]}}{\alpha(1 - 2^{1-n})}.$$

Clearly from (2.7.23) it follows that $\tilde{\theta}_2$ is unbiased for θ_2 and $\tilde{\sigma}_2$ is unbiased for σ_2. By using (2.7.24), (2.7.25) and (2.7.26), we get the variances of $\tilde{\theta}_2$ and $\tilde{\sigma}_2$ as,

$$\text{Var}[\tilde{\theta}_2] = \frac{\pi^2}{3}\sigma_2^2$$

and

$$\text{Var}[\tilde{\sigma}_2] = \frac{1}{\alpha^2(1-2^{1-n})^2}\Big[(\frac{2\pi^2}{3}) - \alpha^2(1-2^{1-n})^2 - 2\alpha^2\{4I_3(1,n) - 2I_1(1,n) - 2I_2(1,n) + 1\}\Big].$$

We have obtained the variance of $\tilde{\theta}_2$, the relative efficiency $\frac{\text{Var}(\tilde{\theta}_2)}{\text{Var}(\hat{\theta}_2)}$ of $\hat{\theta}_2$ relative to $\tilde{\theta}_2$ for $n = 2(1)10; \alpha = 0.25(0.25)0.75$ and are provided in table 2.7.2. Again we have obtained the variance of $\tilde{\sigma}_2$, the relative efficiency $\frac{\text{Var}(\tilde{\sigma}_2)}{\text{Var}(\hat{\sigma}_2)}$ of $\hat{\sigma}_2$ relative to $\tilde{\sigma}_2$ for $n = 2(1)10;\ \alpha = 0.25(0.25)0.75$ are provided in table 2.7.3.

Remark 2.7.1. We can see that the BLUE $\hat{\theta}_2$ of θ_2 does not depend much on the association parameter α but the BLUE $\hat{\sigma}_2$ of σ_2 depends very much on α and our assumption is that α is known. Therefore in the situation where α is unknown we introduce a rough estimator for α as follows, in order to make our estimators $\hat{\theta}_2$ and $\hat{\sigma}_2$ useful for the α unknown situation.

For *MTBLD* the correlation coefficient between the two variates is given by $\rho = \frac{3}{\pi^2}\alpha$. If r is the simple correlation coefficient between R_i and $R_{[i]}, i = 1, 2, 3, \cdots$ then a rough moment type estimator for α is obtained by equating r with the population correlation coefficient ρ and is obtained as,

$$\hat{\alpha} = \begin{cases} -1, & \text{if} \quad r \le -\frac{3}{\pi^2} \\ 1, & \text{if} \quad r \ge \frac{3}{\pi^2} \\ r\frac{\pi^2}{3}, & \text{otherwise.} \end{cases}$$

Remark 2.7.2. From the tables we can see that the efficiency of the BLUE of θ_2, the location parameter ranges from 1 to 1.25 and the efficiency of the BLUE of σ_2 the scale parameter ranges from 1 to 1.75. It is clear that the efficiency of the BLUE of σ_2 is better than the efficiency of the BLUE of θ_2. However, one should keep in mind that competitors are naive estimators because those are the only available estimators to obtain the relative efficiency of our estimators in this situation.

References

Ahsanullah, M. and Nevzorov, V. B. (2000). Some distributions of induced record values, *Biometrical Journal*, **42**, 1069-1081.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*, Wiley, New York.

Balakrishnan, N., Ahsanullah M. and Chan, P. S. (1995). On the logistic record values and associated inference, *Journal of Applied Statistical Science*, **2**, 233-248.

Balasubramanian, K. and Beg, M. I. (1997). Concomitants of order statistics in Morgenstern type bivariate exponential distribution, *Journal of Applied Statistical Science*, **5**, 233-245.

Barnett, Y., Green, P.J. and Robinson, A. (1976). Concomitants and correlation estimates, *Biometrika*, **63**, 323-328.

Begum, A. A. and Khan, A. H. (2000a). Concomitants of order statistics from Marshall and Olkin's bivariate Weibull distribution, *Calcutta Statistical Association Bulletin*, **50**, 65-70.

Begum, A. A. and Khan, A. H. (2000b). Concomitants of order statistics from Gumbel's bivariate logistic distribution, *Journal of the Indian Society for Probability and Statistics*, **5**, 51-64.

Bhattacharya, P.K. (1974). Convergence of sample paths of normalized sums of induced order statistics, *Annals of Statistics*, **2**, 1034-1039.

Castillo, E. (1988). *Extreme Value Theory in Engineering*, Academic Press, New York.

Chacko, M. and Thomas, P. Y. (2004). Estimation of a parameter of Morgenstern type bivariate uniform distribution based on concomitants of order statistics and concomitants of record values, *Journal of the Kerala Statistical Association*, **15**, 13-26.

Chacko, M. and Thomas, P. Y. (2005). Concomitants of record values arising from Morgenstern type bivariate logistic distribution and some of their applications in parameter estimation, *Metrika* (to appear).

Coles, S.G. and Tawn, A. J. (1994). Statistical methods for multivariate extremes: An application to structural design, *Applied Statistics*, **43**, 1-48.

David, H.A. and Nagaraja, H. N. (2003). *Order Statistics*, 3rd edition, John Wiley, New York.

Gill, P.S., Tiku, M.L. and Vaughan, D.C. (1990). Inference problems in life testing under multivariate normality, *Journal of Applied Statistics*, **17**, 133-149.

Harrel, F.E. and Sen., P.K. (1979). Statistical inference for censored bivariate normal distributions based on induced order statistics, *Biometrika*, **66**, 293-298.

Houchens, R. L. (1984). *Record Value Theory and Inference*, Ph. D. Dissertation, University of California, Riverside, California.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, John Wiley, New York.

Scaria, J. and Nair, N. U. (1999). On concomitants of order statistics from Morgenstern family, *Biometrical Journal*, **41**, 483-489.

Spruill, N.L. and Gastwirth, J. (1982). On the estimation of correlation coefficient from grouped data, *Journal of American Statistical Association*, **77**, 614-620.

Yang, S.S. (1977). General distribution theory of the concomitants of order statistics, *Annals of Statistics*, **5**, 996-1002.

Yeo, W.B. and David, H.A. (1984). Selection through an associated characteristic with application to the random effects model, *Journal of American Statistical Association*, **79**, 399-405.