CHAPTER 4

BIRTH AND DEATH PROCESSES AND ORDER STATISTICS

[*The first four sections are based on the lectures of Professor P.R. Parthasarathy of the Department of Mathematics, Indian Institute of Technology, Madras, Chennai 600 036, India, at the* 5^{th} *SERC School. These sections will be devoted to birth and death processes and the remaining sections will be devoted to order statistics based on the lectures of Dr. P. Yageen Thomas, Department of Statistics, University of Kerala, Trivandrum, Kerala, India*]

4.1. Birth and Death Processes

4.1.0. Introduction

An important sub-class of Markov chains with continuous time parameter space is birth and death processes (BDPs), whose state space is the non-negative integers.These processes are characterized by the property that if a transition occurs, then this transition leads to a neighboring state.

The description of the process is as follows: The process sojourns in a given state i for a random length of time following an exponential distribution with parameter $\lambda_i + \mu_i$. When leaving state i, the process enters state $(i + 1)$ or state $(i - 1)$. The motion is analogous to that of a random walk except that transitions occur at random times rather than fixed time periods.

BDPs are frequently used as models of the growth of biological populations. A remarkable variety of dynamic behavior exhibited by many species of plants, insects and animals has stimulated great interest in the development of both biological experiments and mathematical models. In many ecological problems such as animal and cell populations, epidemics, plant tissues, learning processes, competition between species, growth patterns are influenced by population size. Such populations do not increase indefinitely, but are limited by, for example, lack of food, cannibalism and overcrowding. There are many deterministic models which describe such density-dependent logistic population growth. They effectively represent the development of a tumor, the growth of viral plaques or the population of spread in the theory of urban development.

Queuing theory is an important application area of BDPs. It has proved to be useful in a wide range of disciplines from computer networks and telecommunications to chemical kinetics and epidemiology. Suppose that customers arrive at a single server facility in accordance with a Poisson process. That is, times between successive arrivals are independent exponential variables having mean $1/\lambda$. Upon arrival, each customer goes directly into service if the server is free, and if not, the customer joins the queue. When the server finishes serving a customer, the customer leaves the system and the next customer in line, if there are any waiting, enters the service. The successive service times are assumed to be independent exponential variables having mean $1/\mu$. The number of persons in the system at time t is a BDP with birth rate λ and death rate μ independent of the number of customers present in the system at that time.

The presence of only one of the components, viz., birth or death also leads to important applications. For example, in the theory of radioactive transformation the radioactive atoms are unstable and disintegrate stochastically. Each of these new atoms is also unstable. By the emission of radioactive particles, these new atoms pass through a number of physical states with specified decay rates from one state to the adjacent state. Thus the radioactive transformations can be modeled as a birth process. Consider the enzyme reaction of blood clotting. In closing a cut, the gelation process of blood clotting is caused by an enzyme known as fibrin, which is formed by fibrogen. This conversion of fibrogen molecules into fibrin molecules follows a birth process.

BDPs on a finite state space cover a large spectrum of operations research and biological systems. Queuing models with finite state space have applications in production and inventory problems, for example, to optimize the size of the storage space, to determine the trade-off between throughput and inventory (or waiting time) and to exhibit the propagation of blockage. The performance of the produce-to-stock manufacturing facility can be obtained from the performance of the finite queuing systems. Network of queues with finite buffers occur widely in computer and telecommunications systems. The phenomena of blocking, starvation and server breakdowns can severely restrict throughput and response time.

There is a steady increase in the applications of BDPs to natural sciences and practical problems including epidemics, queues and inventories and reliability, production management, computer-communication systems, neutron propagation, optics, chemical reactions, construction and mining and compartmental models.

Therefore, the broad field of applications of birth and death models amply justifies an intensive study of BDPs.

4.1.1. Transient analysis

In the analysis of BDPs, the emphasis is often laid on the steady state solutions while the transient or time-dependent analysis has received less attention. The assumptions required to derive the steady state solutions to queuing systems are seldom satisfied in the design and analysis of real systems. Also, steady state measures of system performance simply do not make sense for systems that never approach equilibrium. Moreover, the steady state results are inappropriate in situations wherein the time horizon of operations is finite. Hence, in many applications, the practitioner needs a knowledge of the time-dependent behavior of the system rather than easily obtainable steady state results. Further, transient solutions are available for a wide class of problems and contribute to a more finely tuned analysis of the costs and benefits of the systems.

BDPs have a huge mathematical literature discussing effective and interesting techniques to determine numerous important quantities like system size probabilities, their stationary behavior, first passage times, etc. The transition probabilities of finite BDPs, for example, can be expressed in terms of sums of exponentials. Its complete unified theory can be used for reference in general Markov chains and other stochastic models.

Recurrence relations play an important role in the transient analysis of BDPs. There is hardly a computational task which does not rely on recursive techniques at one time or another. The widespread use of recurrence relations can be ascribed to their intrinsic constructive quality and the great ease with which they are amenable to mechanization. In particular, " three-term recurrence relations " form the nucleus of continued fractions (CFs), orthogonal polynomials (OPs) and BDPs.

The study of the time dependent behavior of BDPs has given rise to many intricate and interesting OPs . Several interesting OPs occur in the study of BDPs with linear and quadratic birth and death rates. This orthogonal representation leads to the spectral measure for BDPs which is important in the study of the transient behavior.

4.2. Birth and Death Processes

Consider a continuous time Markov chain $\{x(t), t \geq 0\}$ with state space $\mathcal{S} = \{0, 1, 2, \dots\}$ with stationary transition probabilities $p_{ij}(t)$, that is,

$$p_{ij}(t) = P\{x(t+s) = j \mid x(s) = i\}.$$

In addition we assume that the $p_{ij}(t)$ satisfy the following postulates:

1. $p_{i,i+1}(h) = \lambda_i h + o(h)$, as $h \downarrow 0$, $i \geq 0$.
2. $p_{i,i-1}(h) = \mu_i h + o(h)$, as $h \downarrow 0$, $i \geq 1$.
3. $p_{ii}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$, as $h \downarrow 0$, $i \geq 0$.
4. $p_{ij}(h) = \delta_{ij}$.
5. $\mu_0 = 0,\ \lambda_0 > 0,\ \mu_i,\ \lambda_i > 0,\ i = 1, 2, \dots.$

The $o(h)$ in each case may depend on i. The matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0\dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0\dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2\dots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3)\dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{4.2.1}$$

is called the *infinitesimal generator* of the process. The parameters λ_i and μ_i are called, respectively, the infinitesimal birth and death rates. The process $x(t)$ is known as a birth and death process. In Postulates 1 and 2 we are assuming that if the process starts in state i, then in a small interval of time the probabilities of the population increasing or decreasing by 1 are essentially proportional to the length of the interval. Sometimes a transition from zero to some ignored state is allowed.

Since the $p_{ij}(t)$ are probabilities, we have $p_{ij}(t) \geq 0$ and

$$\sum_{j=0}^{\infty} p_{ij}(t) = 1.$$

In order to obtain the probability that $x(t) = n$, we must specify where the process starts or more generally the probability distribution for the initial state. We then have

$$P\{x(t) = n\} = \sum_{i=0}^{\infty} P\{x(0) = i\} p_{in}(t).$$

4.2.1. Waiting times

With the aid of the above assumptions we may calculate the distribution of the random variable t_i which is the waiting time of $x(t)$ in state i; that is, given the process in state i, we are interested in calculating the distribution of the time t_i until it first leaves state i. Letting

$$P\{t_i \geq t\} = G_i(t)$$

it follows easily by the Markov property that as $h \downarrow 0$

$$G_i(t + h) = G_i(t)G_i(h) = G_i(t)(p_{ii}(h) + o(h)) = G_i(t)[1 - (\lambda_i + \mu_i)h] + o(h)$$

or

$$\frac{G_i(t + h) - G_i(t)}{h} = -(\lambda_i + \mu_i)G_i(t) + o(1),$$

so that

$$G_i'(t) = -(\lambda_i + \mu_i)G_i(t).$$

If we use the condition $G_i(0) = 1$, the solution of this equation is

$$G_i(t) = \mathrm{e}^{-(\lambda_i + \mu_i)t}.$$

That is, t_i follows an exponential distribution with mean $(\lambda_i + \mu_i)^{-1}$.

According to Postulates 1 and 2, during a time duration of length h a transition occurs from state i to $i + 1$ with probability $\lambda_i h + o(h)$ and from state i to $i - 1$ with probability $\mu_i h + o(h)$. It follows intuitively that, given that a transition occurs at time t, the probability this transition is to state $i + 1$ is $\lambda_i(\lambda_i + \mu_i)^{-1}$ and to state $i - 1$ is $\mu_i(\lambda_i + \mu_i)^{-1}$.

The description of the motion of $x(t)$ is as follows: The process sojourns in a given state i for a random length of time following an exponential distribution with parameter $\lambda_i + \mu_i$. When leaving state i the process enters either state $i + 1$ or state $i - 1$ with probabilities $\dfrac{\lambda_i}{\lambda_i + \mu_i}$, or, $\dfrac{\mu_i}{\lambda_i + \mu_i}$, respectively. The motion is analogous to that of a random walk except that transitions occur at random times rather than at fixed time periods.

4.2.2. The Kolmogorov equations

The transition probabilities $p_{ij}(t)$ satisfy a system of differential equations known as the *backward Kolmogorov differential equations*. These are given by

$$\begin{aligned} p'_{0j}(t) &= -\lambda_0 p_{0j}(t) + \lambda_0 p_{1j}(t), \\ p'_{ij}(t) &= \mu_1 p_{i-1,j}(t) - (\lambda_i + \mu_i) p_{ij}(t) + \lambda_{i+1,j}(t), \qquad i \geq 1, \end{aligned} \tag{4.2.2}$$

and the boundary condition $p_{ij}(0) = \delta_{ij}$.

To derive these we have from (4.2.1)

$$\begin{aligned} p_{ij}(t+h) &= \sum_{k=0}^{\infty} p_{ik}(h) p_{kj}(t) = p_{i,i-1}(h) p_{i-1,j}(t) + \\ &\qquad p_{ii}(h) p_{ij}(t) + p_{i,i+1} p_{i+1,j}(t) + \sum_k p_{ik}(h) p_{kj}(t), \end{aligned} \tag{4.2.3}$$

where the last summation is over all $k \neq i-1, i, i+1$. Using Postulates 1, 2, and 3 we obtain

$$\begin{aligned} \sum_k p_{ik}(h) p_{kj}(t) &\leq \sum_k p_{ik}(h) \\ &= 1 - [p_{ii}(h) + p_{i,i-1}(h) + p_{i,i+1}(h)] \\ &= 1 - [1 - (\lambda_i + \mu_i)h + o(h) + \mu_i h + o(h) + \lambda_i h + o(h)] \\ &= o(h), \end{aligned}$$

so that

$$p_{ij}(t+h) = \mu_i h p_{i-1,j}(t) + (1 - (\lambda_i + \mu_i)h) p_{ij}(t) + \lambda_i h p_{i+1,j}(t) + o(h).$$

Transposing the terms $p_{ij}(t)$ to the left-hand side and dividing the equation by h, we obtain the equations (4.2.2), after letting $h \downarrow 0$.

The backward equations are deduced by decomposing the time interval $(0, t+h)$, where h is positive and small, into the two periods

$$(0, h), \qquad (h, t+h),$$

and examining the transitions in each period separately. The equations (4.2.2) feature the initial state as the variable. A different result arises from splitting the time interval $(0, t + h)$ into the two periods

$$(0, t), \qquad (t, t + h)$$

and adapting the preceding analysis. In this viewpoint, under more stringent conditions, we can derive a further system of differential equations

$$\begin{aligned} p'_{i0}(t) &= -\lambda_0 p_{i0}(t) + \mu_1 p_{i1}(t), \\ p'_{ij}(t) &= \lambda_{j-1} p_{i,j-1}(t) - (\lambda_j + \mu_j) p_{ij}(t) + \mu_{j+1} p_{i,j+1}(t), \qquad j \geq 1, \end{aligned} \tag{4.2.4}$$

with the same initial condition $p_{ij}(0) = \delta_{ij}$. These are known as the *forward Kolmogorov differential equations*. To do this we interchange t and h in (4.2.3) and under stronger assumptions in addition to Postulates 1, 2, and 3 it can be shown that the last term is again $o(h)$. The remainder of the argument is the same as before.

In general the infinitesimal parameters $\{\lambda_n, \mu_n\}$ may not determine a unique stochastic process obeying

$$\sum_{j=0}^{\infty} p_{mn}(t) = 1 \text{ and } p_{mn}(t + s) = \sum_{k=0}^{\infty} p_{mn}(t) p_{kn}(s).$$

A sufficient condition that there exists a unique Markov process with $p_{mn}(t)$ satisfying the above is that

$$\sum_{n=0}^{\infty} \pi_n \sum_{k=0}^{n} \frac{1}{\lambda_n \pi_n} = \infty,$$

where $\pi_0 = 1$ and

$$\pi_n = \pi_{n-1} \frac{\lambda_{n-1}}{\mu_n} = \frac{\lambda_0 \ldots \lambda_{n-1}}{\mu_1 \ldots \mu_n}, \; n \in \mathcal{S} \setminus \{0\}. \tag{4.2.5}$$

The quantities π_n are called the *potential coefficients*.

We discuss now briefly the behavior of $p_{ij}(t)$ as t becomes large. It can be proved that the limits

$$\lim_{t \to \infty} p_{ij}(t) = p_j \tag{4.2.6}$$

exist and are independent of the initial state i and also that they satisfy the equations

$$0 = -\lambda_0 p_0 + \mu_1 p_1 \tag{4.2.7}$$

$$0 = \lambda_{n-1} p_{n-1} - (\lambda_n + \mu_n) p_n + \mu_{n+1} p_{n+1}, \ n \geq 1. \tag{4.2.8}$$

which is obtained by setting the left-hand side of (4.2.4) equal to zero. The convergence of $\sum_j p_j$ follows since $\sum_j p_{ij}(t) = 1$. If $\sum_j p_j = 1$, then the sequence $\{p_j\}$ is called a *stationary distribution* (steady state probabilities). The reason for this is that p_j also satisfy

$$p_j = \sum_{i=0}^{\infty} p_i p_{ij}(t), \tag{4.2.9}$$

which tells us that if the process starts in state i with probability p_i, then at any given time t it will be in state i with same probability p_i. The proof of (4.2.9) follows from (4.2.1) and (4.2.6) and if we let $t \uparrow \infty$ and use the fact that $\sum_{i=0}^{\infty} p_i < \infty$. The solution to (4.2.7) and (4.2.8) is obtained by rewriting them as

$$\begin{aligned} \mu_{n+1} p_{n+1} - \lambda_n p_n &= \mu_n p_n - \lambda_{n-1} p_{n-1} \\ &= \cdots = \mu_1 p_1 - \lambda_0 p_0 = 0. \end{aligned}$$

Therefore,

$$p_j = \pi_j p_0, \quad j = 1, 2, \ldots . \tag{4.2.10}$$

In order that the sequence $\{p_j\}$ define a distribution we must have $\sum_{j=0}^{\infty} p_j = 1$. If $\sum \pi_k < \infty$, we see in this case that

$$p_j = \frac{\pi_j}{\sum_{k=0}^{\infty} \pi_k}, \qquad j = 0, 1, 2, \ldots .$$

If $\sum \pi_k = \infty$, then necessarily $p_0 = 0$ and the p_j are all zero. Hence, we do have a limiting stationary distribution.

Example 4.2.1. *Poisson process*: For this process, the rates are given by $\lambda_n = \lambda$ (constant) and $\mu_n = 0$. This process plays an important role as an arrival process in many contexts. In this case, the value of j in the function f_j of the random variable technique is 1 and

$$f_1(x(t)) = \lambda, \qquad \text{independent of } x(t).$$

The probability generating function $P(s, t)$ is given by

$$\frac{\partial p}{\partial t} = \lambda(s-1)P, \ P(s, 0) = s$$

which leads to $P(s,t) = \mathrm{e}^{\lambda(s-1)t}$. Therefore,

$$\begin{aligned} P_n(t) &= \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots \\ E[x(t)] &= \lambda t \quad \text{and } \mathrm{Var}[x(t)] = \lambda t, \quad t \geq 0. \end{aligned}$$

For a Poisson process, the time between two subsequent occurrences of events is exponentially distributed with parameter λ. That is, if t_1 denotes the time of the first event and t_n denotes the time between the $(n-1)^{th}$ and n^{th} events, the random variables $\{t_i,\ i = 1, 2, \ldots\}$ are independent exponentially distributed with the parameter of the Poisson process.

Example 4.2.2. *Simple birth process (Yule process)*: For this process, the rates are given by $\lambda_n = n\lambda$ and $\mu_n = 0$. This process models a population in which each member acts independently and gives birth at an exponential rate λ. The simplest example of a pure birth process is the Poisson process, which has a constant birth rate $\lambda_n = \lambda,\ n \geq 0$. Here

$$f_1(x(t)) = \lambda x(t).$$

Starting with a single individual, the probability generating function using random variable technique $P(s,t)$ can be given by

$$\frac{\partial P}{\partial t} = \lambda s(s-1)\frac{\partial P}{\partial s},\ P(s,0) = s.$$

Then,

$$P(s,t) = \left\{1 - \mathrm{e}^{-\lambda t}\left(1 - \frac{1}{s}\right)\right\}^{-1}.$$

Therefore, $\{x(t)\}$ follows a geometric distribution. Hence,

$$E[x(t)] = \mathrm{e}^{\lambda t} \text{ and } \mathrm{Var}[x(t)] = \mathrm{e}^{\lambda t}(\mathrm{e}^{\lambda t} - 1).$$

Example 4.2.3. The rates are given by

$$\lambda_n = (N-n)\lambda, \quad \mu_{n+1} = (n+1)\mu, \quad n = 0, 1, \ldots, N-1.$$

Thus,

$$\begin{aligned} P_n(t) = {} & \frac{\mu^m}{(\lambda+\mu)^N}\left[1 - \mathrm{e}^{-(\lambda+\mu)t}\right]^m \left[\mu + \lambda\mathrm{e}^{-(\lambda+\mu)t}\right]^{N-m} \\ & \times \sum_{i=0}^{n} \binom{m}{i}\binom{N-m}{n-i}\left[\frac{\lambda + \mu\mathrm{e}^{-(\lambda+\mu)t}}{\mu(1 - \mathrm{e}^{-(\lambda+\mu)t})}\right]^i \left[\frac{\lambda(1 - \mathrm{e}^{-(\lambda+\mu)t})}{\mu + \lambda\mathrm{e}^{-(\lambda+\mu)t}}\right]^{n-i}, \end{aligned}$$

where $x(0) = m$. The mean and variance of $x(t)$ at time t are given by

$$\begin{aligned}
E[x(t)] &= \frac{1}{\lambda+\mu}\left[\lambda N - e^{-(\lambda+\mu)t}(\lambda N - m(\lambda+\mu))\right], \\
\mathrm{Var}[x(t)] &= \frac{1}{(\lambda+\mu)^2}\left[\lambda\mu N - e^{-(\lambda+\mu)}(\lambda N(\mu-\lambda) - m(\mu^2-\lambda^2)^2)\right. \\
&\quad \left. - e^{-2(\lambda+\mu)t}(\lambda^2 N + m(\mu^2-\lambda^2))\right].
\end{aligned}$$

Exercises 4.2.

4.2.1. If the birth and death rates are given by $\lambda_n = \lambda$, $\mu_n = n\mu$. Obtain the probability generating function of the number of customers at time t.($M/M/\infty$, Infinite server queue)

4.2.2. For $\lambda_n = \lambda$, $\mu_n = n\mu$, obtain $P_{00}(t)$. ($M/M/1$ queueing system)

4.2.3. For a BDP on {0,1} obtain the transition probabilities.

4.2.4. For a BDP with rates $\lambda_0 = 1$ and other $\lambda_n = 1/2$ and $\mu_n = 1/2$, show that $P_{00}(t) = e^{-t} I_0(t)$.

4.2.5. Arrivals to a supermarket is a Poisson process with arrival rate 10 per hour. Given that 100 persons arrived up to 1 PM, what is the probability that 2 more persons will arrive in 10 minutes? What is the average number of arrivals between 11 AM and 1 PM?

4.2.6. Consider a Poisson process. Given that an event has occurred before T find the pdf of this arrival.

4.2.7. Arrivals at a telephone booth are considered to be Poisson, with an average time of 10 minutes between one arrival and the next. The length of a phone call is assumed to be distributed exponentially with mean 3 minutes. Find the probability that a person arriving at the booth will have to wait and his average time spent in the booth.

4.3. Continued Fractions

A CF is denoted by

$$\cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}$$

or economically by

$$\frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + \frac{a_3|}{|b_3} + \cdots, \tag{4.3.1}$$

where a_n and b_n are real or complex numbers. This fraction can be terminated by retaining the terms $a_1,\ b_1,\ a_2,\ b_2, \ldots, a_n,\ b_n$ and dropping all the remaining terms $a_{n+1},\ b_{n+1}, \ldots$. The number obtained by this operation is called the n^{th} convergent or n^{th} approximant and is denoted by A_n/B_n. Both A_n and B_n satisfy the recurrence relation

$$U_n = a_n U_{n-2} + b_n U_{n-1} \tag{4.3.2}$$

with initial values $A_0 = 0,\ A_1 = a_1$ and $B_0 = 1,\ B_1 = b_1$. $\frac{A_{2n}}{B_{2n}}$ ($\frac{A_{2n+1}}{B_{2n+1}}$) is called the even (odd) part of the CF.

A CF is denoted by

$$\cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}$$

or economically by

$$\frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + \frac{a_3|}{|b_3} + \cdots, \tag{4.3.3}$$

where a_n and b_n are real or complex numbers. This fraction can be terminated by retaining the terms $a_1,\ b_1,\ a_2,\ b_2, \ldots, a_n,\ b_n$ and dropping all the remaining terms $a_{n+1},\ b_{n+1}, \ldots$. The number obtained by this operation is called the n^{th} convergent or n^{th} approximant and is denoted by A_n/B_n. Both A_n and B_n satisfy the recurrence relation

$$U_n = a_n U_{n-2} + b_n U_{n-1} \tag{4.3.4}$$

with initial values $A_0 = 0,\ A_1 = a_1$ and $B_0 = 1,\ B_1 = b_1$. $\frac{A_{2n}}{B_{2n}}$ $(\frac{A_{2n+1}}{B_{2n+1}})$ is called the even (odd) part of the CF.

A CF is said to be convergent if at most a finite number of its denominators B_n vanish and the limit of its sequence of approximants

$$\lim_{n\to\infty} \frac{A_n}{B_n}, \tag{4.3.5}$$

exists and is finite. Otherwise, the CF is said to be divergent. If

$$a_n > 0\ ,\ b_n > 0\ ,\ \sum_{n=1}^{\infty} \sqrt{\frac{b_n b_{n-1}}{a_{n+1}}} = \infty,$$

then the CF is convergent. The value of a CF is defined to be the limit (4.3.5). No value is assigned to a divergent CF.

For any convergent CF, the exact value of the fraction is between any two neighboring convergents. All even numbered convergents lie to the left of the exact value, that is, they give an approximation to the exact value by defect. All odd numbered convergents lie to the right of the exact value, that is, they give an approximation to the exact value by excess. For a terminating CF, that is, a fraction with finite number of terms, the last convergent coincides with the exact value. If a CF is non-terminating, the sequence of convergents is infinite. For example,

$$\sqrt{2} = 1 + (\sqrt{2} - 1) = 1 + \frac{1}{1 + \sqrt{2}} = 1 + \frac{1|}{|2} + \frac{1|}{|2} + \cdots .$$

The successive convergents are

$$\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \ldots .$$

That is,

$$1.0,\ 1.5,\ 1.4,\ 1.417,\ 1.4138,\ 1.41429,\ 1.41420, \ldots .$$

Observe that

$$\frac{7}{5} < \sqrt{2} < \frac{3}{2},\ \frac{41}{29} < \sqrt{2} < \frac{99}{70},\ \text{ and so on.}$$

Using the power series expansion of

$$\frac{8\tan\left(\dfrac{\pi h}{2}\right) - \tan(\pi h)}{3h} = \pi - \frac{\pi^5}{480}h^4 - \frac{5\pi^7}{16\cdot 7!}h^6 - \cdots ,$$

we get a CF representation of π as follows:

$$\pi = 3 + \frac{1|}{|7} + \frac{1|}{|15} + \frac{1|}{|1} + \frac{1|}{|292} + \frac{1|}{|1} + \cdots .$$

Its successive convergents are $\frac{3}{1}, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \cdots$ We observe that

$$\left|\pi - \frac{355}{113}\right| < 3 \times 10^{-7}.$$

This shows how fast the CF is converging to the exact value. This result is spectacular. However,

$$e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}} dt = \frac{1|}{|x} + \frac{1|}{|x} + \frac{2|}{|x} + \frac{3|}{|x} + \cdots$$

requires about 70 terms to pin the error down to 2.4×10^{-7} when $x = 1$.

It is well-known that a sequence of OPs $\{P_n(x)\}$ satisfy a three-term recurrence relation

$$P_{n+1}(x) = (x + b_n)P_n(x) - a_n P_{n-1}(x) \ , n \geq 0 \tag{4.3.6}$$

with $P_0 = 1$, $P_{-1} = 0$ and $a_n > 0$. This recurrence relation when compared with the recurrence relation (4.3.4) suggests consideration of the Jacobi CF

$$\frac{1|}{|x + b_1} - \frac{a_1|}{|x + b_2} - \frac{a_2|}{|x + b_3} - \cdots .$$

In this case the n^{th} convergent is a rational function $A_{n-1}(x)/B_n(x)$ whose denominator $B_n(x)$ is a polynomial of degree n and numerator $A_{n-1}(x)$ is a polynomial of degree $n - 1$. The sequences $\{B_n(x)\}$ and $\{A_n(x)\}$ satisfy (4.3.6) with $B_0(x) = 1$, $B_1(x) = x + b_1$ and $A_0(x) = 0$, $A_1(x) = 1$. Further, $A_{n-1}(x)$ form an associated OP system and the zeros of $A_{n-1}(x)$ and $B_n(x)$ are distinct and interlace.

4.3.1. CFs and BDPs

In this section, approximate transient system size probability values for an infinite BDP $\{x(t), t \geq 0\}$ is obtained using the CF technique discussed below. Define

$$L_{r-1} = \prod_{k=0}^{r-1} \lambda_k; \qquad M_r = \prod_{k=1}^{r} \mu_k, \ r = 0, 1, 2, \ldots$$

with $L_{-1} = 1 = M_0$. Define

$$f_r(s) = (-1)^r M_r \int_0^\infty \mathrm{e}^{-st} P_r(t)\mathrm{d}t, \qquad r = 0, 1, 2, \ldots .$$

Taking the Laplace transform of the system of equations given by (4.2.4) and assuming $m = 0$, $f_0(s)$ gives the expression

$$f_0(s) = \frac{1}{s + \lambda_0 + \dfrac{f_1(s)}{f_0(s)}}.$$

Similarly,

$$\frac{f_r(s)}{f_{r-1}(s)} = \frac{-\lambda_{r-1}\mu_r}{s + \lambda_r + \mu_r + \dfrac{f_{r+1}(s)}{f_r(s)}}, \; r = 1, 2, 3, \ldots .$$

This leads to

$$f_0(s) = \frac{1 \rfloor}{\lfloor s + \lambda_0} - \frac{\lambda_0\mu_1 \rfloor}{\lfloor s + \lambda_1 + \mu_1} - \frac{\lambda_1\mu_2 \rfloor}{\lfloor s + \lambda_2 + \mu_2} \cdots . \tag{4.3.7}$$

Let the n^{th} approximant of $f_0(s)$ be given by $\dfrac{A_{n-1}(s)}{B_n(s)}$ where

$$\begin{aligned}
A_0(s) &= 1 \\
A_1(s) &= s + \lambda_1 + \mu_1 \\
A_n(s) &= (s + \lambda_{n-1} + \mu_{n-1})A_{n-1}(s) - \lambda_{n-2}\mu_{n-1}A_{n-2}(s), \; n = 2, 4, \ldots,
\end{aligned}$$

and

$$\begin{aligned}
B_1(s) &= s + \lambda_0 \\
B_2(s) &= (s + \lambda_1 + \mu_1)B_1(s) - \lambda_0\mu_1 \\
B_n(s) &= (s + \lambda_{n-1} + \mu_{n-1})B_{n-1}(s) - \lambda_{n-2}\mu_{n-1}B_{n-2}(s), \; n = 3, 4, \ldots .
\end{aligned}$$

Then

$$f_0(s) \approx \frac{A_{n-1}(s)}{B_n(s)}, \tag{4.3.8}$$

where $\approx$ signifies "approximately" and $B_n(s)$ can be written in tridiagonal determinant form as follows:-

$$B_n(s) = \begin{vmatrix} s+\lambda_0 & 1 & & & & \\ \lambda_0\mu_1 & s+\lambda_1+\mu_1 & 1 & & & \\ & \lambda_1\mu_2 & s+\lambda_2+\mu_2 & 1 & & \\ & & & \ddots & & \\ & & & & \lambda_{n-2}\mu_{n-1} & s+\lambda_{n-1}+\mu_{n-1} \end{vmatrix}_{n\times n}. \quad (4.3.9)$$

$A_{n-1}(s)$ is obtained from $B_n(s)$ by deleting the first row and first column.

We observe that $B_n(s)$ given by (4.3.9) is clearly zero when $-s$ is an eigenvalue of the matrix

$$F_n = \begin{bmatrix} \lambda_0 & \mu_1 & & & & \\ \lambda_0 & \lambda_1+\mu_1 & \mu_2 & & & \\ & \lambda_1 & \lambda_2+\mu_2 & \mu_3 & & \\ & & & \ddots & & \\ & & & & \lambda_{n-2} & \lambda_{n-1}+\mu_{n-1} \end{bmatrix}_{n\times n}.$$

This matrix is quasi-symmetric and can be transformed into a real symmetric matrix E_n by a similarity transformation $E_n := D_n^{-1}F_nD_n$, where

$$D_n := \mathrm{Diag}\left\{1,\ \sqrt{L_0M_1},\ \sqrt{L_1M_2}, \ldots,\ \sqrt{L_{n-2}M_{n-1}}\right\}.$$

The matrix thus obtained is given by

$$E_n = \begin{bmatrix} \lambda_0 & \sqrt{\lambda_0\mu_1} & & & & \\ \sqrt{\lambda_0\mu_1} & \lambda_1+\mu_1 & \sqrt{\lambda_1\mu_2} & & & \\ & \sqrt{\lambda_1\mu_2} & \lambda_2+\mu_2 & \sqrt{\lambda_2\mu_3} & & \\ & & & \ddots & & \\ & & & & \sqrt{\lambda_{n-2}\mu_{n-1}} & \lambda_{n-1}+\mu_{n-1} \end{bmatrix}_{n\times n}.$$

This is a real symmetric diagonal dominant positive definite tridiagonal matrix with non-zero subdiagonal elements and therefore the eigenvalues are real and distinct. We denote these eigenvalues by $s_1^{(n)}, s_2^{(n)}, s_3^{(n)}, \ldots, s_n^{(n)}$, that is, $-s_1^{(n)}, -s_2^{(n)}, s_3^{(n)}, \ldots, -s_n^{(n)}$ are the roots of $B_n(s)$. Similarly, the roots of $A_{n-1}(s)$ are negative, real and distinct and let $-z_1^{(n)}, -z_2^{(n)}, -z_3^{(n)}, \ldots, -z_{n-1}^{(n)}$ be its roots.

Using partial fractions, (4.3.8) can be expressed as

$$f_0(s) \approx \sum_{j=1}^{n} \frac{\prod_{r=1}^{n-1}(z_r^{(n)} - s_j^{(n)})}{(s + s_j^{(n)}) \prod_{r=1, r\neq j}^{n} (s_r^{(n)} - s_j^{(n)})}.$$

On inverting we get,

$$P_0(t) \approx \sum_{j=1}^{n} \frac{\prod_{r=1}^{n-1}(z_r^{(n)} - s_j^{(n)})}{\prod_{r=1, r\neq j}^{n} (s_r^{(n)} - s_j^{(n)})} e^{-s_j^{(n)} t}.$$

Example 4.3.1. *BDP with constant birth and death rates*: The rates are given by

$$\lambda_n = \lambda; \qquad \mu_{n+1} = \mu, \quad n = 0, 1, 2, \ldots, N-1. \tag{4.3.10}$$

This process corresponds to an $M/M/1/N$ queuing model with arrival rate λ and service rate μ. We get the roots of s_k of $B_{N+1}(s)$ as follows:

$$s_0 = 0, \; s_k = -\lambda - \mu + 2\sqrt{\lambda\mu} \cos \frac{k\pi}{N+1}, \qquad k = 1, 2, ..., N.$$

By induction, for $i = 0, 1, 2, \ldots, N$,

$$B_i(-s_k) = \begin{cases} \lambda^i, & \text{if } k = 0 \\ 4^i (\lambda\mu)^{\frac{i}{2}} y_{i,k} - 4^{i-1}\mu (\lambda\mu)^{\frac{i-1}{2}} y_{i-1,k}, & \text{if } k = 1, 2, 3, \ldots, N. \end{cases},$$

where, for $k = 1, 2, 3, \ldots, N$ and $i = 2, 3, 4, \ldots, N+1$,

$$y_{i-1,k} = \prod_{j=1}^{i} \sin\left(\frac{[(i+1)k + (N+1)j]\,\pi}{2(N+1)(i+1)}\right) \sin\left(\frac{[(i+1)k - (N+1)j]\,\pi}{2(N+1)(i+1)}\right).$$

We get for $n = 0, 1, 2, \ldots, N$,

$$P_n(t) = \frac{\lambda^{N-n}\mu^n}{\prod_{j=1}^{N}\left\{\lambda + \mu - 2\sqrt{\lambda\mu}\, cos\frac{j\pi}{N+1}\right\}}$$

$$+4^{n+m-2}\mu\sqrt{\rho}\sum_{k=1}^{N}\frac{\{4\sqrt{\rho}Y_{n,k} - Y_{n-1,k}\}\{4\sqrt{\rho}Y_{m,k} - y_{m-1,k}\}\, y_{N-1,k}\mathrm{e}^{-t(\lambda+\mu-2\sqrt{\lambda\mu}\cos\frac{k\pi}{N+1})}}{\left(1 + \rho - 2\sqrt{\rho}\, cos\frac{k\pi}{N+1}\right)\prod_{j=1, j\neq k}^{N}\sin\left(\frac{(j+k)\pi}{2(N+1)}\right)\sin\left(\frac{(j-k)\pi}{2(N+1)}\right)},$$

where $\rho = \frac{\lambda}{\mu}$ and m (≥ 0) is the initial state.

Example 4.3.2.

$$\text{Rates} \quad : \quad \lambda_{n-1} = \frac{(\gamma+n-1)(\beta+n)}{(\gamma+2n-2)(\gamma+2n-1)},$$

$$\mu_n = \frac{n(\gamma+n-1-\beta)}{(\gamma+2n-1)(\gamma+2n)},\ n = 1, 2, 3, \ldots$$

$$\text{CF} \quad : \quad z\,{}_2F_1(\beta+1, 1; \gamma+1, -z) = \frac{z|}{|1} + \frac{\lambda_0 z|}{|1} + \frac{\mu_1 z|}{|1} + \frac{\lambda_1 z|}{|1} + \frac{\mu_2 z|}{|1} + \quad \cdots$$

$$P_{00}(t) \quad = \quad \sum_{k=0}^{\infty}\frac{(-1)^k(\beta+1)_k}{(\gamma+1)_k}\frac{t^k}{k!} \ = \ {}_1F_1(\beta+1, \gamma+1, -t).$$

In particular, if $\beta = 0$ and $\gamma = 1$, then

$$P_{00}(t) = \frac{1-\mathrm{e}^{-t}}{t}.$$

Example 4.3.3.

Rates : $\lambda_{n-1} = (N-n+1)p\ ,\ \mu_n = nq\ ,\ n = 1, 2, \cdots, N$

$$\text{CF} : \sum_{j=0}^{N}\frac{b(N, j; p)}{z+j} = \frac{1|}{|z} + \frac{Np|}{|1} + \frac{q|}{|z} + \frac{(N-1)p|}{|1} + \frac{2q|}{|z} + \cdots + \frac{Nq|}{|z}, \qquad \Re(z) > 0$$

where $b(N, j; p) := \binom{N}{j}p^j q^{N-j}\ ,\ j = 0, 1, 2, \cdots, N\ ,\ q = 1-p$

$$P_{00}(t) = \sum_{j=0}^{N} b(N, j; p)\mathrm{e}^{-jt}.$$

Example 4.3.4.

$$\begin{aligned} \text{Rates} \quad &: \quad \lambda_{n-1} = \frac{1}{2n},\ \mu_n = \frac{1}{2n+1},\ n = 1,2,3,\cdots \\ \text{CF} \quad &: \quad {}_1F_1(1; z+1; z) = 1 + \frac{2z|}{|2} + \frac{3z|}{|3} + \frac{4z|}{|4} + \frac{5z|}{|5} + \cdots \end{aligned}$$

for any complex number z,

$$\begin{aligned} P_{00}(t) &= \sum_{m=1}^{\infty} \frac{e^{-m} m^{m-2}}{(m-1)!} e^{-t/m}. \\ P_{00}(0) &= 1 \end{aligned}$$

Exercises 4.3.

4.3.1. If the rates are $\lambda_{n-1} = \frac{\alpha U_n(1/\alpha)}{2U_{n-1}(1/\alpha)},\ \mu_n = \frac{\alpha U_{n-1}(1/\alpha)}{2U_n(1/\alpha)},\ \alpha < 1,\ n = 1,2,3,\cdots$, show that $P_{00}(t) = 2e^{-t}\frac{I_1(\alpha t)}{\alpha t}$.

4.3.2. For the rates

$$\lambda_{n-1} = \frac{nr}{1-r},\ \mu_n = \frac{n}{1-r},\ r < 1, n = 1,2,3,\cdots,$$

prove that $P_{00}(t) = \frac{1-r}{1-re^{-t}}$.

4.3.3. For a BDP with rates $\lambda_n = 1/2$ and $\mu_n = n/(2n+1)$, show that $P_{00}(t) = \frac{1-e^{-t}}{t}$.

4.3.4. Obtain a CF expansion of $\sqrt{5}$

4.3.5. Express $\frac{3551}{1132}$ as a CF and calculate the first four convergents.

4.3.6. Use recurrence relation to evaluate the tridiagonal determinant with $2cos\theta$ in the diagonal and 1 in the subdiagonal and superdiagonal.

4.3.7. Give one CF expansion of π.

4.4. Orthogonal Polynomials

Definition 4.4.1. A sequence of polynomials $\{Q_n(x)\}$, where $Q_n(x)$ is of exact degree n in x, is said to be orthogonal with respect to a Lebesgue-Stieltjes measure $\mathrm{d}\alpha(x)$ if

$$\int_{-\infty}^{\infty} Q_m(x)Q_n(x)\mathrm{d}\alpha(x) = 0, \quad m \neq n. \tag{4.4.1}$$

Implicit in this definition is the assumption that the moments

$$m_n = \int_{-\infty}^{\infty} x^n \mathrm{d}\alpha(x), \quad n = 0, 1, 2, \dots \tag{4.4.2}$$

exist and are finite. If the non-decreasing, real-valued, bounded function $\alpha(x)$ also happens to be absolutely continuous with $\mathrm{d}\alpha(x) = w(x)\mathrm{d}x$, $w(x) \geq 0$, then (4.4.1) reduces to

$$\int_{-\infty}^{\infty} Q_m(x)Q_n(x)w(x)\mathrm{d}x = 0, \quad m \neq n, \tag{4.4.3}$$

and the sequence $\{Q_n(x)\}$ is said to be orthogonal with respect to the weight function $w(x)$. If, on the other hand, $\alpha(x)$ is a step-function with jumps w_j at $x = x_j$, $j = 0, 1, 2, \dots$, then (4.4.1) takes the form of a sum:

$$\sum_{j=0}^{\infty} Q_m(x_j)Q_n(x_j)w_j = 0, \quad m \neq n. \tag{4.4.4}$$

In this case we refer to the sequence $\{Q_n(x)\}$ as OPs of a discrete variable.

General OPs are characterized by the fact that they satisfy a three-term recurrence relation of the form

$$xQ_n(x) = a_n Q_{n+1}(x) + b_n Q_n(x) + c_n Q_{n-1}(x), \tag{4.4.5}$$

where $Q_{-1}(x) = 0$, $Q_0(x) = 1$, a_n, b_n, c_n are real and $a_n c_{n+1} > 0$ for $n = 0, 1, 2, \dots$. This is an important characterization which has been the starting point for much of the modern work on general OPs. Monic OPs for a positive Borel measure α on the real line are polynomials P_n, $n = 0, 1, 2, \dots$, of degree n and leading coefficient one such that

$$\int P_n(x)x^k \mathrm{d}\alpha(x) = 0, \quad k = 0, 1, 2, \dots, n-1. \tag{4.4.6}$$

These n orthogonality conditions give n linear equations for the n unknown coefficients $a_{k,n},\ k = 0, 1, 2, \ldots, n-1$, of the monic polynomial $P_n(x) = \sum_{k=0}^{n} a_{kn} x^k$, where $a_{nn} = 1$. This system of n equations for n unknowns always has a unique solution since the matrix with entries $\int x^{i+j} \mathrm{d}\alpha(x),\ 0 \le i, j \le n-1$, known as *Gram matrix*, is positive definite and hence nonsingular. It is well known that such polynomials satisfy a three-term recurrence relation

$$P_{n+1}(x) = (x - b_n)P_n(x) - a_n^2 P_{n-1}(x), \quad n \ge 0 \tag{4.4.7}$$

with $P_0 = 1$ and $P_{-1} = 0$. Often it is more convenient to consider the orthonormal polynomials $p_n(x) = \gamma_n P_n(x)$ for which

$$\int p_n(x) p_m(x) \mathrm{d}\alpha(x) = \delta_{mn}, \quad m, m \ge 0 \tag{4.4.8}$$

so that $\gamma_n = \left(\int P_n^2(x) \mathrm{d}\alpha(x)\right)^{-1/2} > 0$. The recurrence relation for these orthonormal polynomials is given by

$$x p_n(x) = a_{n+1} p_{n+1}(x) + b_n p_n(x) + a_n p_{n-1}(x), \quad n \ge 0, \tag{4.4.9}$$

where

$$a_n = \int x p_{n-1}(x) p_n(x) \mathrm{d}\alpha(x) = \frac{\gamma_{n-1}}{\gamma_n}, \quad b_n = \int x p_n^2(x) \mathrm{d}\alpha(x). \tag{4.4.10}$$

Let us mention at this point the best known examples: the classical OPs of Jacobi, Laguerre and Hermite. These three families of OPs satisfy a second order linear differential equation, their derivatives are again OPs of the same family but with different parameters.

Apart from the three-term recurrence relations, these OPs satisfy a number of relationships of the same general form. They satisfy the second order differential equation

$$g_2(x) p_n'' + g_1(x) p_1' + d_n p_n = 0.$$

Between any two zeros of $p_n(x)$ there is a zero of $p_{n+1}(x)$. Many interesting properties of these polynomials depend on their connection with problems of interpolation and mechanical quadrature.

Remarkably for many specific BDPs of practical interest, the weight function $\mathrm{d}\alpha(x)$ has been explicitly determined and general formulas are derived for the OPs. The spectral representation then gives considerable insight into the time-dependent behavior of the BDPs.

For each system of OPs R_n satisfying the recurrence relation

$$-xR_n(x) = A_nR_{n+1}(x) - (A_n + C_n)R_n(x) + C_nR_{n-1}(x),$$

we can introduce an extra scaling parameter $\tau \neq 0$ and study the polynomials $S_n(x) = R_n(\tau x)$, which satisfy the recurrence relation

$$-xS_n(x) = \frac{A_n}{\tau}S_{n+1}(x) - \frac{(A_n + C_n)}{\tau}S_n(x) + \frac{C_n}{\tau}S_{n-1}(x).$$

If the polynomials R_n are orthogonal on the finite set $\{x_0, x_1, \dots, x_N\}$, then the polynomials S_n are orthogonal on the set $\{x_0/\tau, x_1/\tau, \dots, x_N/\tau\}$

Example 4.4.1. *Chebyshev polynomials of the second kind*: Chebyshev polynomials of the second kind U_n satisfy the recurrence relation

$$2xU_n(x) = U_{n+1}(x) + U_{n-1}(x), \qquad n \geq 1, \tag{4.4.11}$$

with $U_0(x) = 1$ and $U_1(x) = 2x$. They are orthogonal on the interval $[-1, 1]$ with respect to the weight $\sqrt{1-x^2}$.

If we consider only the finite set $\{U_0, U_1, \dots, U_N\}$, then this is also orthogonal on the set $\{\cos j\pi/(N+2),\ j = 1, 2, \dots, N+1\}$, i.e., the roots of U_{N+1}. Consider orthonormal polynomials p_n satisfying the relation

$$xp_n(x) = a_{n+1}p_{n+1}(x) + b_np_n(x) + a_np_{n-1}(x),$$

with $p_0(x) = 1$ and $p_{-1}(x) = 0$. Then the Chebyshev polynomials of the second kind are the orthonormal polynomials with constant recurrence coefficients $a_n = 1/2$ $(n = 1, 2, \dots)$ and $b_n = 0$ $(n = 0, 1, 2, \dots)$. If we change the first recurrence coefficient to $b_0 = b/2$, then we get polynomials P_n which are a special case of co-recursive OPs. Obviously U_n and U_{n-1} are both solutions of the recurrence relation

$$2xP_n(x) = P_{n+1}(x) + P_{n-1}(x),$$

hence the general solution of this second order linear recurrence is $P_n(x) = AU_n(x) + BU_{n-1}(x)$, where A and B are determined from the initial conditions $P_0(x) = 1$ and $P_1(x) = 2x - b$. This gives

$$P_n(x) = U_n(x) - bU_{n-1}(x).$$

If we require a finite system of OPs, then we take $a_{N+1} = 0$. Finally taking $b_N = 1/2b$ gives

$$\begin{aligned} P_{N+1}(x) &= \left(x - \frac{1}{2b}\right)P_N(x) - \frac{1}{2}P_{N-1}(x) \\ &= xU_N(x) - bxU_{N-1}(x) - \frac{1}{2b}U_N(x) + \frac{b}{2}U_{N-2}(x). \end{aligned}$$

Now use $U_{N-2}(x) = 2xU_{N-1}(x) - U_N(x)$ to find

$$P_{N+1}(x) = \left(x - \frac{b+1/b}{2}\right) U_N(x). \tag{4.4.12}$$

Hence this finite system of modified Chebyshev polynomials, with $b_0 = b/2$ and $b_N = 1/2b$, is orthogonal on the zeros of P_{N+1}, i.e., on

$$\left\{ \frac{b+1/b}{2}, \cos\frac{j\pi}{N+1}, j = 1, 2, \ldots, N \right\}.$$

We note that the Chebyshev polynomials of the second kind can be written as

$$U_n(\cos(\theta)) \quad = \quad \frac{\sin((n+1)\theta)}{\sin(\theta)}, \; n = 0, 1, \ldots \; ; \; 0 \le \theta \le \pi. \tag{4.4.13}$$

The modified Chebyshev polynomials of second kind $V_n(x)$ and $W_n(x)$ are defined as follows for $n = 0, 1, \ldots$:

$$V_n(\cos(\theta)) \quad = \quad \frac{\cos\left((n+\frac{1}{2})\theta\right)}{\cos\left(\frac{\theta}{2}\right)} \tag{4.4.14}$$

and

$$W_n(\cos(\theta)) \quad = \quad \frac{\sin\left((n+\frac{1}{2})\theta\right)}{\sin(\frac{\theta}{2})}. \tag{4.4.15}$$

The significance of these observations is that the properties of sines and cosines can be used to establish many of the properties of Chebyshev polynomials. Chebyshev polynomials have acquired great practical importance in polynomial approximation methods. Specifically, it has been shown that a series of Chebyshev polynomials converges much more rapidly than a power series.

Chebyshev polynomials satisfy the generating function

$$(1 - 2xt + t^2)^{-1} = \sum_{n=0}^{\infty} U_n(x) t^n.$$

We can write (4.4.13) as

$$U_n(x) = \sum_{k=0}^{[n/2]} \binom{n-k}{k} (-1)^k (2x)^{n-2k}.$$

Define

$$V_n(x) = U_n(x) - U_{n-1}(x) \text{ and } W_n(x) = U_n(x) + U_{n-1}(x), \; n \ge 1,$$

where $V_0(x) = W_0(x) = 1$. The polynomials $V_n(x)$ and $W_n(x)$ are orthogonal in $(-1, 1)$ with weight functions $\sqrt{\frac{1+x}{1-x}}$ and $\sqrt{\frac{1-x}{1+x}}$, respectively (Szegö (1959)).

Example 4.4.2. *Krawtchouk Polynomials*: These polynomials $K_n(x; p, N)$, with $0 < p < 1$, satisfy the recurrence relation

$$-xK_n(x) = p(N-n)K_{n+1}(x) - [p(N-n) + n(1-p)]K_n(x) + n(1-p)K_{n-1}(x). \quad (4.4.16)$$

They are orthogonal on the set $\{0, 1, 2, \dots, N\}$ with respect to the binomial distribution with parameters N and p.

Example 4.4.3. *Dual Hahn Polynomials $H_n(x; \gamma, \delta, N)$*: These polynomials satisfy the recurrence relation

$$-xH_n(x) = A_n H_{n+1}(x) - (A_n + C_n)H_n(x) + C_n H_{n-1}(x), \quad (4.4.17)$$

where $A_n = (N-n)(n+d+1),\ C_n = n(N-n+\delta+1)$.
They are orthogonal on the set $\{n(n+d+\delta+1),\ n = 0, 1, 2, \dots, N\}$.

Example 4.4.4. *Racah Polynomials $R_n(x; \alpha, \beta, \gamma, \delta)$*: These polynomials with $\alpha + 1 = -N$ or $\beta + \delta + 1 = -N$ or $\gamma + 1 = -N$, satisfy the recurrence relation

$$-xR_n(x) = A_n R_{n+1}(x) - (A_n + C_n)R_n(x) + C_n R_{n-1}(x), \quad (4.4.18)$$

where

$$A_n = -\frac{(n+\alpha+\beta+1)(n+\alpha+1)(n+\beta+\delta+1)(n+d+1)}{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)},$$

$$C_n = -\frac{n(n+\beta)(n+\alpha+\beta-d)(n+\alpha-\delta)}{(2n+\alpha+\beta)(2n+\alpha+\beta+1)}.$$

They are orthogonal on the set $\{n(n+d+\delta+1),\ n = 0, 1, 2, \dots, N\}$.

4.4.1. OPs and BDPs

For every BDP $\{x(t), t \geq 0\}$ with state space $\mathcal{S}$, one has Karlin-McGregor polynomials $Q_n(x)$ $(n = 0, 1, \dots)$, satisfying the three-term recurrence relation

$$\begin{gathered} Q_{-1}(x) := 0, \quad Q_0(x) := 1 \\ \lambda_n Q_{n+1}(x) + (x - \lambda_n - \mu_n)Q_n(x) + \mu_n Q_{n-1}(x) = 0, \qquad n \geq 0 \end{gathered} \quad (4.4.19)$$

which may be written in matrix notation as

$$Q_0(x) = 1, \quad -x\boldsymbol{Q}(x) = \mathbf{Q}^{\top}\mathcal{Q}(x), \quad (4.4.20)$$

where $\boldsymbol{Q}(x)$ is the column vector with elements $Q_i(s),\ i \in \mathcal{S}$ and

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_0 & \mu_1 & 0 & & \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4.4.21}$$

We observe that the monic polynomials $P_n(x) := (-1)^n(\lambda_0\lambda_1 \ldots \lambda_n)Q_n(x),\ n > 0$, satisfy the recurrence formula

$$\begin{aligned} xP_n(x) &= P_{n+1}(x) + (\lambda_n + \mu_n)P_n(x) + \lambda_{n-1}\mu_n P_{n-1}(x), \quad n > 0, \\ P_0(x) &= 1, \quad P_1(x) = x - \lambda_0 \end{aligned} \tag{4.4.22}$$

and that $P_n(-x)$ can be interpreted as the characteristic polynomials of the matrix $\boldsymbol{Q}^T$ truncated at the n^{th} row and column. Since $\lambda_{n-1}\mu_n > 0$ for $n > 0$, the recurrence relation (4.4.22) and hence (4.4.19) defines a sequence of polynomials which is orthogonal with respect to a positive measure. Moreover, the parameters in the recurrence relation (4.4.22) are such that $\{P_n(x)\}$ is orthogonal with respect to a measure on the nonnegative real axis. Let $q_n(x)$ denote the n^{th} orthonormal polynomial corresponding to $Q_n(x)$ or $P_n(x)$. It can easily be verified that

$$q_n(x) = \pi_n^{1/2} Q_n(x), \quad n \geq 0, \tag{4.4.23}$$

where π_n are given by (4.2.5).

We now obtain the transition probability functions $P_n(t)$, which satisfy (4.2.4), in terms of the polynomials $Q_n(x)$ for the given rates λ_n and μ_n. Suppose that the polynomials $\{q_n(x)\}$ are orthonormal with respect to the measure $d\alpha$, with support in $[0, \infty)$, that is,

$$\int_0^\infty q_n(x)q_m(x)\mathrm{d}\alpha(x) = \delta_{m,n}$$

and define the functions

$$f_i(x,t) = \sum_{j=0}^\infty P_j(t)Q_j(x) = \sum_{j=0}^\infty \pi_j^{-1/2} P_j(t)q_j(x), \quad i \in \mathcal{S} \tag{4.4.24}$$

or equivalently

$$f(x,t) = \boldsymbol{P}(t)\boldsymbol{Q}(x), \tag{4.4.25}$$

where $\boldsymbol{P}(t) = (P_{mn}(t))_{m,n=0}^\infty$. We observe that this matrix satisfies

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q}^\top \tag{4.4.26}$$

with

$$\boldsymbol{P}(0) = \boldsymbol{I}, \tag{4.4.27}$$

where $\boldsymbol{I} = (\delta_{ij})_{i,j=0}^{\infty}$ is the infinite identity matrix. Using (4.4.20) and (4.4.26), we note that (4.4.25) satisfies the equation

$$\frac{\partial}{\partial t} f(x,t) = \boldsymbol{P}'(t)\boldsymbol{Q}(x) = \boldsymbol{P}(t)\boldsymbol{Q}^{\top}\boldsymbol{Q}(x) = -xf(x,t)$$

and by (4.4.27), the initial condition is

$$f(x,0) = \boldsymbol{Q}(x).$$

Hence,

$$f(x,t) = \mathrm{e}^{-xt}\boldsymbol{Q}(x)$$

or

$$f_i(x,t) = \mathrm{e}^{-xt}Q_i(x), \quad i \in \mathcal{S}.$$

By (4.4.24), $\pi_j^{-1/2}P_j(t)$ is the j^{th} Fourier coefficient of $f_i(x,t)$ with respect to the orthonormal system $\{q_n(x)\}$ and hence

$$\pi_j^{-1/2}P_j(t) = \int_0^{\infty} \mathrm{e}^{-xt}Q_i(x)q_j(x)\mathrm{d}\alpha(x), \quad i, j \in \mathcal{S}$$

or

$$P_j(t) \;=\; \pi_j \int_0^{\infty} \mathrm{e}^{-xt}Q_i(x)Q_j(x)\mathrm{d}\alpha(x), \quad i, j \in \mathcal{S}. \tag{4.4.28}$$

For every finite BDP $\{x(t), t \geq 0\}_f$ with finite state space $\mathcal{S}_f$ one has Karlin-McGregor polynomials Q_n, $n = 0, 1, 2, \ldots, N$, defined by the recurrence relation

$$-xQ_n(x) = \lambda_n Q_{n+1}(x) - (\lambda_n + \mu_n)Q_n(x) + \mu_n Q_{n-1}(x), \;\; n = 1, 2, \ldots, N-1 \tag{4.4.29}$$

with initial values

$$Q_0(x) = 1, \quad Q_1(x) = \frac{\lambda_0 - x}{\lambda_0}$$

and the polynomial Q_{N+1} of degree $N+1$, defined by

$$-xQ_N(x) = Q_{N+1}(x) - \mu_N Q_N(x) + \mu_N Q_{N-1}(x). \tag{4.4.30}$$

Karlin and McGregor have shown that Q_{N+1} has $N+1$ distinct, real zeros $s_0 < s_1 < s_2 < \cdots < s_N$. Furthermore, the polynomials $\{Q_n,\ n = 0, 1, \ldots, N\}$ are orthogonal on the set $\{s_0, s_1, \ldots, s_N\}$:

$$\sum_{k=0}^{N} Q_n(s_k)Q_m(s_k)\rho_k = 0, \qquad 0 \le n \ne m \le N, \tag{4.4.31}$$

where ρ_k are positive weights given by

$$\rho_k := \frac{1}{\sum_{i=0}^{N} Q_i^2(s_k)\pi_i} > 0. \tag{4.4.32}$$

They gave the spectral representation of the $P_n(t)$ in terms of these polynomials, as

$$P_n(t) = \pi_n \sum_{k=0}^{N} e^{-s_k t} Q_m(s_k)Q_n(s_k)\rho_k, \qquad n \in \mathcal{S}_f. \tag{4.4.33}$$

The monic orthogonal Karlin-McGregor polynomials, corresponding to the polynomials $Q_n(s)$, are given by $P_n(x) := (-1)^n(\lambda_0\lambda_1 \ldots \lambda_n)Q_n(x),\ n = 0, 1, \ldots, N$ and they satisfy the recurrence formula

$$xP_n(x) = P_{n+1}(x) + (\lambda_n + \mu_n)P_n(x) + \lambda_{n-1}\mu_n P_{n-1}(x), \quad n = 1, 2, \ldots, N \tag{4.4.34}$$

with $P_0(x) = 1$ and $P_{-1}(x) = 0$. So we can also find the monic Karlin-McGregor polynomial $P_{N+1}(x)$, and its zeros are important because the Karlin-McGregor polynomials are orthogonal on the set consisting of these zeros.

Example 4.4.5. Consider the BDP discussed in example 4.3.1. For this BDP, the Karlin-McGregor polynomials $Q_n(x)$ satisfy the recurrence relation

$$-xQ_n(x) = \lambda Q_{n+1}(x) - (\lambda + \mu)Q_n(x) + \mu Q_{n-1}(x), \qquad n = 1, 2, \ldots, N-1, \tag{4.4.35}$$

with $Q_0(x) = 1$ and $Q_1(x) = (\lambda - x)/\lambda$. Take the modifi ed Chebyshev polynomials $P_n(x) = U_n(x) - bU_{n-1}(x)$ with $b = \sqrt{\mu/\lambda}$, then the recurrence relation (4.4.11) and the change of variable

$$Q_n(x) = \left(\frac{\mu}{\lambda}\right)^{\frac{n}{2}} P_n\left(\frac{\lambda + \mu - x}{2\sqrt{\lambda\mu}}\right)$$

give the recurrence (4.4.35) with the initial conditions $Q_0(x) = 1$ and $Q_1(x) = (\lambda - x)/\lambda$. So the Karlin-McGregor polynomials for this fi nite BDP turn out to be the modifi ed Chebyshev polynomials

$$U_n(y) - \sqrt{\mu/\lambda}\, U_{n-1}(y), \qquad y = \frac{\lambda + \mu - x}{2\sqrt{\mu\lambda}},$$

which are orthogonal on the set $\{0, \lambda + \mu - 2\sqrt{\lambda\mu}\cos\frac{k\pi}{N+1},\ k = 1, 2, \ldots, N\}$ (see Example 4.4.1 for details).

Example 4.4.6. Consider the BDP discussed in example 4.4.2.

The Karlin-McGregor polynomials for this process are the Krawtchouk polynomials

$$K_n\left(\frac{x}{\lambda + \mu};\ \frac{\lambda}{\lambda + \mu},\ N\right),$$

which are orthogonal on the set

$$\{0, \lambda + \mu, 2(\lambda + \mu), \ldots, N(\lambda + \mu)\}.$$

Example 4.4.7. Consider a finite BDP with quadratic rates

$$\begin{aligned} \lambda_n &= (N - n)(b - (n - 1)c), \quad n = 0, 1, \ldots, N - 1, \\ \mu_n &= (N + n + 1)(b + (n + 2)c), \quad n = 1, 2, \ldots, N, \end{aligned} \tag{4.4.36}$$

with $b > 0,\ -b/(N + 2) < c < 0$. Conditions on the parameters are to ensure the positivity of the rates.

The Karlin-McGregor polynomials for this process are the Racah polynomials

$$R_n\left(-\frac{x}{4c};\ -\frac{1}{2},\ \frac{1}{2},\ -N - 1,\ -\frac{2b + 5c}{2c}\right),$$

which are orthogonal on the set

$$\left\{-4cn\left(n - \frac{2b + 5c}{2c} - N\right),\ n = 0, 1, 2, \ldots, N\right\}.$$

Example 4.4.8. Consider a finite BDP with quadratic rates

$$\begin{aligned} \lambda_n &= (N - n)(b + (n - 1)d), \quad n = 0, 1, \ldots, N - 1, \\ \mu_n &= n(c - (n - 1)d), \quad n = 1, 2, \ldots, N, \end{aligned} \tag{4.4.37}$$

where $b, c > 0$ and

$$\begin{cases} d \le 0, & \text{if } N = 1 \text{ or } 2, \\ \frac{-b}{N-2} < d \le 0, & \text{if } N > 2. \end{cases}$$

The conditions on the parameters are to ensure the positivity of the rates.

The Karlin-McGregor polynomials for this BDP are the dual Hahn polynomials

$$H_n\left(\frac{x}{d}; \frac{b-2d}{d}, \frac{c-Nd}{d}, N\right),$$

which are orthogonal on the set

$$\left\{dn\left(n + \frac{b+c}{d} - N - 1\right),\ n = 0, 1, 2, \dots, N\right\}.$$

Exercises 4.4.

4.4.1. Identify the birth and death rates associated with the continued fraction $\frac{1}{N}\sum_{j=0}^{N-1}\frac{1}{z+j} = \frac{1|}{|z} + \frac{\lambda_0|}{|1} + \frac{\mu_1|}{|z} + \cdots \frac{\mu_N|}{|z}$, $z > 0$, and show that $P_{00}(t) = \frac{1}{N}\sum_{j=0}^{N-1} e^{-jt}$.

4.4.2. Identify the birth and death rates associated with the continued fraction $\sum_{j=0}^{N}\frac{b(N,j;p)}{z+j} = \frac{1|}{|z} + \frac{Np|}{|1} + \frac{q|}{|z} + \frac{(N-1)p|}{|1} + \frac{2q|}{|z} + \cdots + \frac{Nq|}{|z}$, $\quad \Re(z) > 0$
where $b(N, j; p) := \begin{pmatrix} N \\ j \end{pmatrix} p^j q^{N-j}$, $j = 0, 1, 2, \cdots, N$, $q = 1 - p$ and show that $P_{00}(t) = \sum_{j=0}^{N} b(N, j; p)e^{-jt}$.

Reference

Parthasarathy, P.R. and Lenin, R.B. (2004). *Birth and Death Process (BDP) Models with Applications – Queueing, Communication Systems, Chemical Models, Biological Models: The State-of-the-Art with a Time-Dependent Perspective*, American Series in Mathematical and Management Sciences 51, Syracuse, NY, American Sciences Press, Inc.

4.5. Estimation of a Parameter by Ranked-set Sampling

[*This section is based on the lectures of P. Yageen Thomas, Department of Statistics, University of Kerala, Trivandrum 695 581, India.*]

4.5.1. Introduction

The concept of ranked-set sampling (RSS) was first introduced by McIntyre (1952) as a process of improving the precision of the sample mean as an estimator of the population mean. ranked-set sampling as described in McIntyre (1952) is applicable whenever ranking of a set of sampling units can be done easily by a judgement method (for a detailed discussion on the theory and applications of ranked-set sampling see, Chen et al (2004). Ranking by judgement method is not recommendable if the judgement method is too crude and is not powerful for ranking by discriminating the units of a moderately large sample. In certain situations, one may prefer exact measurement of some easily measurable variable associated with the study variable rather than ranking the units by a crude judgement method. Suppose the variable of interest say Y, is difficult or much expensive to measure, but an auxiliary variable X correlated with Y is readily measurable and can be ordered exactly. In this case as an alternative to McIntyre (1952) method of ranked-set sampling, Stokes (1977) used an auxiliary variable for the ranking of the sampling units. The procedure of ranked-set sampling described by Stokes (1977) using auxiliary variate is as follows: Choose n^2 independent units, arrange them randomly into n sets each with n units and observe the value of the auxiliary variable X on each of these units. In the first set, that unit for which the measurement on the auxiliary variable is the smallest is chosen. In the second set, that unit for which the measurement on the auxiliary variable is the second smallest is chosen. The procedure is repeated until in the last set, that unit for which the measurement on the auxiliary variable is the largest is chosen. The resulting new set of n units chosen by one from each set as described above is called the RSS defined by Stokes (1977). If $X_{(r)r}$ is the observation measured on the auxiliary variable X from the unit chosen from the rth set then we write $Y_{[r]r}$ to denote the corresponding measurement made on the study variable Y on this unit, then $Y_{[r]r}$, $r = 1, 2, \ldots, n$ form the ranked-set sample. Clearly $Y_{[r]r}$ is the concomitant of the rth order statistic arising from the rth sample.

A striking example for the application of the ranked-set sampling as proposed by Stokes (1977) is given in Bain (1978, p.99), where the study variate Y represents the oil pollution of sea water and the auxiliary variable X represents the tar deposit in the nearby sea shore. Clearly collecting sea water sample and measuring the oil pollution in it is strenuous and expensive. However the prevalence of pollution in the sea water is much reflected by the tar deposit in the surrounding terminal sea shore. In this example ranking the pollution level of sea water based on the tar deposit in the sea shore is more natural and scientific than ranking it visually or by judgement method.

Stokes (1995) has considered the estimation of parameters of location-scale family of distributions using RSS. Lam et al (1994, 1996) have obtained the BLUEs of location and scale parameters of exponential distribution and logistic distribution. The Fisher information contained in RSS have been discussed by Chen (2000) and Chen and Bai (2000). Stokes (1980) has considered the method of estimation of correlation coefficient of bivariate normal distribution by using RSS. Modarres and Zheng (2004) have considered the problem of estimation of dependence parameter using RSS. Robust estimate of correlation coefficient for bivariate normal distribution have been developed by Zheng and Modarres (2006). Stokes (1977) has suggested the ranked-set sample mean as an estimator for the mean of the study variate Y, when an auxiliary variable X is used for ranking the sample units, under the assumption that (X, Y) follows a bivariate normal distribution. Barnett and Moore (1997) have improved the estimator of Stokes (1977) by deriving the Best Linear Unbiased Estimator (BLUE) of the mean of the study variate Y, based on ranked set sample obtained on the study variate Y.

In this paper we are trying to estimate the mean of the population, under a situation where in measurement of observations are strenuous and expensive. Bain (1978, P.99) has proposed an exponential distribution for the study variate Y, the oil pollution of the sea samples. Thus in this paper we assume a Morgenstern type bivariate exponential distribution (MTBED) corresponding to a bivariate random variable (X, Y), where X denotes the auxiliary variable (such as tar deposit in the sea shore) and Y denotes the study variable (such as the oil pollution in the sea water). A random variable (X, Y) follows MTBED if its probability density function (pdf) is given by (see, Kotz et al, 2000, P.353)

$$f(x,y)=\begin{cases}\frac{1}{\theta_1\theta_2}\exp\left\{-\frac{x}{\theta_1}-\frac{y}{\theta_2}\right\} & \left[1+\alpha\left(1-2\exp\left\{-\frac{x}{\theta_1}\right\}\right)\left(1-2\exp\left\{-\frac{y}{\theta_2}\right\}\right)\right],\\ & x>0, y>0; -1\le\alpha\le 1; \theta_1>0, \theta_2>0\\ 0, \text{ otherwise.}\end{cases} \tag{4.5.1}$$

In Section 4.5 of this note we have derived an estimator θ_2^* of the parameter θ_2 involved in (4.5.1) using the ranked-set sample mean. It may be noted that if (X, Y) has a MTBED as defined in (4.5.1) then the marginal distributions of both X and Y have exponential distributions and the pdf of Y is given by

$$f_Y(y)=\frac{1}{\theta_2}e^{-y/\theta_2}, y>0, \theta_2>0. \tag{4.5.2}$$

The Cramer-Rao Lower Bound (CRLB) of any unbiased estimator of θ_2 based on a random sample of size n drawn from (4.5.2) is $\frac{\theta_2^2}{n}$. In Section 4.5, we have also shown that the variance of the proposed estimator θ_2^* is strictly less than the CRLB $\frac{\theta_2^2}{n}$ (associated with the marginal pdf (4.5.2)) for all $\alpha\in A$ where $A=[-1,1]-\{0\}$. In this section, we have further made an efficiency comparison between θ_2^* and the maximum likelihood estimator (MLE) $\tilde{\theta}_2$based on a random sample of size n arising from (4.5.1). In Section 4.5, we have derived the BLUE $\hat{\theta}_2$ of θ_2 involved in MTBED based on the ranked-set sample and obtained the efficiency of $\hat{\theta}_2$ relative to $\tilde{\theta}_2$. In Section 4.5, we have considered the situation where we apply censoring and ranking on each sample and ultimately used a ranked-set sample arising out of this procedure to estimate θ_2. In Section 4.5, we have derived the BLUE of θ_2 involved in (4.5.1) using unbalanced multistage ranked-set sampling method. In this section, we have further analysed the efficiency of the estimator of θ_2 based on unbalanced MSRSS when compared with the $\tilde{\theta}_2$ the MLE of θ_2.

4.5.2. ranked-set sample mean as an estimator

Let (X, Y) be a bivariate random variable which follows a MTBED with pdf defined by (4.5.1). Suppose RSS in the sense of stokes (1977) as explained in Section 4.5 is carried out. Let $X_{(r)r}$ be the observation measured on the auxiliary variate X in the rth unit of the RSS and let $Y_{[r]r}$ be the measurement made on the Y variate of the same unit, $r=1,2,\cdots,n$. Then clearly $Y_{[r]r}$ is distributed as the concomitant of *rth* order statistic of a random sample of n arising from (4.5.1). By using the expressions for means and variances of concomitants of order statistics arising from MTBED obtained by Scaria and Nair (1999), the mean and variance of $Y_{[r]r}$ for $1 \le r \le n$ are given below

$$E[Y_{[r]r}] = \theta_2\left[1 - \alpha\frac{n-2r+1}{2(n+1)}\right], \tag{4.5.3}$$

$$\mathrm{Var}[Y_{[r]r}] = \theta_2^2\left[1 - \alpha\frac{n-2r+1}{2(n+1)} - \alpha^2\frac{(n-2r+1)^2}{4(n+1)^2}\right]. \tag{4.5.4}$$

Since $Y_{[r]r}$ and $Y_{[s]s}$ for $r \neq s$ are measurements on Y made from two units involved in two independent samples we have

$$\mathrm{Cov}[Y_{[r]r}, Y_{[s]s}] = 0, r \neq s. \tag{4.5.5}$$

In the following theorem we propose an estimator θ_2^* of θ_2 involved in (4.5.1) and prove that it is an unbiased estimator of θ_2.

Theorem 4.5.1. *Let $Y_{[r]r}$, $r = 1, 2, \cdots, n$ be the ranked set sample observations on a study variate Y obtained out of ranking made on an auxiliary variate X, when (X, Y) follows MTBED as defined in (4.5.1). Then the ranked-set sample mean given by*

$$\theta_2^* = \frac{1}{n}\sum_{r=1}^{n} Y_{[r]r},$$

is an unbiased estimator of θ_2 and its variance is given by,

$$\mathrm{Var}[\theta_2^*] = \frac{\theta_2^2}{n}\left[1 - \frac{\alpha^2}{4n}\sum_{r=1}^{n}\left(\frac{n-2r+1}{n+1}\right)^2\right].$$

Proof 4.5.1.

$$E[\theta_2^*] = \frac{1}{n}\sum_{r=1}^{n} E[Y_{[r]r}].$$

On using (4.5.3) for $E[Y_{[r]r}]$ in the above equation we get,

$$E[\theta_2^*] = \frac{1}{n}\sum_{r=1}^{n}[1 - \alpha\frac{n-2r+1}{2(n+1)}]\theta_2 \tag{4.5.6}$$

It is clear to note that

$$\sum_{r=1}^{n}(n-2r+1) = 0. \tag{4.5.7}$$

Applying (4.5.7) in (4.5.6) we get,

$$E[\theta_2^*] = \theta_2.$$

Thus θ_2^* is an unbiased estimtor of θ_2. The variance of θ_2^* is given by

$$\text{Var}[\theta_2^*] = \frac{1}{n^2}\sum_{r=1}^{n}\text{Var}(Y_{[r]r}).$$

Now using (4.5.4) and (4.5.7) in the above sum we get,

$$\text{Var}[\theta_2^*] = \frac{\theta_2^2}{n}\left[1 - \frac{\alpha^2}{4n}\sum_{r=1}^{n}\left(\frac{n-2r+1}{n+1}\right)^2\right].$$

Thus the theorem is proved.

Now we compare the variance of θ_2^* with the CRLB θ_2^2/n of any unbiased estimator of θ_2 involved in (4.5.2) which is the marginal distribution of Y in (4.5.1). If we write $e_1(\theta_2^*)$ to denote the ratio of θ_2^2/n with $\text{Var}(\theta_2^*)$ then we have,

$$e_1(\theta_2^*) = \frac{1}{\left[1 - \frac{\alpha^2}{4n}\sum_{r=1}^{n}\left(\frac{n-2r+1}{n+1}\right)^2\right]}. \tag{4.5.8}$$

It is very trivial to note that

$$e_1(\theta_2^*) \geq 1.$$

Thus we conclude that there is some gain in efficiency on the estimator θ_2^* due to ranked-set sampling. The reason for the above conclusion is that a ranked-set sample always provides more information than simple random sample even if ranking is imperfect (see, Chen et al, 2004, P. 58). It is to be noted that $\text{Var}(\theta_2^*)$ is a decreasing function of α^2 and hence the gain in efficiency of the estimator θ_2^* increases as $|\alpha|$ increases.

Again on simplifying (4.5.8) we get

$$e_1(\theta_2^*) = \frac{1}{1 - \frac{\alpha^2}{4}\left[\frac{2}{3}\left(\frac{2+1/n}{1+1/n}\right) - 1\right]}.$$

Then

$$\lim_{n\to\infty} e_1(\theta_2^*) = \lim_{n\to\infty} \frac{1}{1-\frac{\alpha^2}{4}[\frac{2}{3}(\frac{2+1/n}{1+1/n})-1]}$$
$$= \frac{1}{1-\frac{\alpha^2}{12}}.$$

From the above relation it is clear that the maximum value for $e_1(\theta_2^*)$ is attained when $|\alpha| = 1$ and in this case $e_1(\theta_2^*)$ tends to 12/11.

Next we obtain the efficiency of θ_2^* by comparing the variance of θ_2^* with the asymptotic variance of MLE of θ_2 involved in MTBED. If (X, Y) follows a MTBED with pdf defined by (4.5.1), then,

$$\frac{\partial \log f(x,y)}{\partial\theta_1} = \frac{1}{\theta_1}\left\{-1+\frac{x}{\theta_1}+\frac{2\alpha x}{\theta_1}\frac{e^{-\frac{x}{\theta_1}}(2e^{-\frac{y}{\theta_2}}-1)}{1+\alpha(2e^{-\frac{x}{\theta_1}}-1)(2e^{-\frac{y}{\theta_2}}-1)}\right\}$$

and

$$\frac{\partial \log f(x,y)}{\partial\theta_2} = \frac{1}{\theta_2}\left\{-1+\frac{y}{\theta_2}+\frac{2\alpha y}{\theta_2}\frac{e^{-\frac{y}{\theta_2}}(2e^{-\frac{x}{\theta_1}}-1)}{1+\alpha(2e^{-\frac{x}{\theta_1}}-1)(2e^{-\frac{y}{\theta_2}}-1)}\right\}.$$

Then we have,

$$I_{\theta_1}(\alpha) = E\left(\frac{\partial \log f(x,y)}{\partial\theta_1}\right)^2,$$
$$= \frac{1}{\theta_1^2}\left\{1+4\alpha^2\int_0^\infty\int_0^\infty \frac{u^2e^{-3u}(2e^{-v}-1)e^{-v}}{\{1+\alpha(2e^{-u}-1)(2e^{-v}-1)\}}dvdu\right\}.$$

$$I_{\theta_2}(\alpha) = E\left(\frac{\partial \log f(x,y)}{\partial\theta_2}\right)^2,$$
$$= \frac{1}{\theta_2^2}\left\{1+4\alpha^2\int_0^\infty\int_0^\infty \frac{v^2e^{-3v}(2e^{-u}-1)e^{-u}}{\{1+\alpha(2e^{-u}-1)(2e^{-v}-1)\}}dvdu\right\}$$

and

$$I_{\theta_1\theta_2}(\alpha) = E\left(\frac{\partial^2 \log f(x,y)}{\partial\theta_1\partial\theta_2}\right),$$
$$= \frac{1}{\theta_1\theta_2}\left\{4\alpha\int_0^\infty\int_0^\infty \frac{uve^{-2u}e^{-2v}}{\{1+\alpha(2e^{-u}-1)(2e^{-v}-1)\}}dvdu\right\}.$$

Thus the Fisher information matrix associated with the random variable (X, Y) is given by

$$I(\alpha) = \begin{pmatrix} I_{\theta_1}(\alpha) & -I_{\theta_1\theta_2}(\alpha) \\ -I_{\theta_1\theta_2}(\alpha) & I_{\theta_2}(\alpha) \end{pmatrix}. \tag{4.5.9}$$

We have evaluated the values of $\theta_1^{-2} I_{\theta_1}(\alpha)$ and $\theta_1^{-1}\theta_2^{-1} I_{\theta_1\theta_2}(\alpha)$ numerically for $\alpha = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ (clearly $\theta_1^{-2} I_{\theta_1}(\alpha) = \theta_2^{-2} I_{\theta_2}(\alpha)$) and are given below.

Table 4.5.1

α	$\theta_1^{-2} I_{\theta_1}(\alpha)$	$\theta_1^{-1}\theta_2^{-1} I_{\theta_1\theta_2}(\alpha)$	α	$\theta_1^{-2} I_{\theta_1}(\alpha)$	$\theta_1^{-1}\theta_2^{-1} I_{\theta_1\theta_2}(\alpha)$
0.25	1.0062	0.0625	-0.25	1.0062	-0.0628
0.50	1.0254	0.1258	-0.50	1.0254	-0.1274
0.75	1.0596	0.1914	-0.75	1.0596	-0.1955
1.00	1.1148	0.2624	-1.00	1.1148	-0.2712

Thus from (4.5.9) the asymptotic variance of the MLE $\tilde{\theta}_2$ of θ_2 involved in MTBED based on a bivariate sample of size n is obtained as

$$\mathrm{Var}(\tilde{\theta}_2) = \frac{1}{n} I_{\theta_2}^{(-1)}(\alpha), \tag{4.5.10}$$

where $I_{\theta_2}^{(-1)}(\alpha)$ is the $(2, 2)th$ element of the inverse of $I(\alpha)$ given by (4.5.9).

We have obtained the efficiency $e(\theta_2^*|\tilde{\theta}_2) = \frac{\mathrm{Var}(\tilde{\theta}_2)}{\mathrm{Var}(\theta_2^*)}$ of θ_2^* relative to $\tilde{\theta}_2$ for $n = 2(2)10(5)20$; $\alpha = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ and are presented in Table 4.5.1. From the table, one can easily see that θ_2^* is more efficient than $\tilde{\theta}_2$ and efficiency increases with n and $|\alpha|$ for $n \geq 4$.

4.5.3. Best linear unbiased estimator

In this section we provide a better estimator of θ_2 than that of θ_2^* by deriving the BLUE $\hat{\theta}_2$ of θ_2 provided the parameter α is known. Let $X_{(r)r}$ be the observation measured on the auxiliary variate X in the rth unit of the RSS and let $Y_{[r]r}$ be the measurement made on the Y variate of the same unit, $r = 1, 2, \cdots, n$. Let

$$\xi_r = 1 - \alpha \frac{n - 2r + 1}{2(n + 1)} \tag{4.5.11}$$

and

$$\delta_r = 1 - \alpha \frac{n-2r+1}{2(n+1)} - \alpha^2 \frac{(n-2r+1)^2}{4(n+1)^2}. \tag{4.5.12}$$

Using (4.5.11) in (4.5.3) and (4.5.12) in (4.5.4) we get $E[Y_{[r]r}] = \theta_2\xi_r, 1 \le r \le n$, and $\mathrm{Var}[Y_{[r]r}] = \theta_2^2\delta_r, 1 \le r \le n$. Clearly from (4.5.5), we have $\mathrm{Cov}[Y_{[r]r}, Y_{[s]s}] = 0,\ r, s = 1, 2, \ldots, n$ *and* $r \ne s$. Let $\mathbf{Y}_{[n]} = (Y_{[1]1}, Y_{[2]2}, \cdots, Y_{[n]n})'$ and if the parameter α involved in ξ_r and δ_r is known then proceeding as in David and Nagaraja (2003, p. 185) the BLUE $\hat{\theta}_2$ of θ_2 is obtained as

$$\hat{\theta}_2 = (\xi' G^{-1} \xi)^{-1} \xi' G^{-1} \mathbf{Y}_{[n]} \tag{4.5.13}$$

and

$$\mathrm{Var}(\hat{\theta}_2) = (\xi' G^{-1} \xi)^{-1} \theta_2^2, \tag{4.5.14}$$

where $\xi = (\xi_1, \xi_2, \cdots, \xi_n)'$ and $G = \mathrm{diag}(\delta_1, \delta_2, \cdots, \delta_n)$. On substituting the values of ξ and G in (4.5.13) and (4.5.14) and simplifying we get

$$\hat{\theta}_2 = \frac{\sum_{r=1}^{n} (\xi_r/\delta_r) Y_{[r]r}}{\sum_{r=1}^{n} \xi_r^2/\delta_r} \tag{4.5.15}$$

and

$$\mathrm{Var}(\hat{\theta}_2) = \frac{1}{\sum_{r=1}^{n} \xi_r^2/\delta_r} \theta_2^2.$$

Thus $\hat{\theta}_2$ as given in (4.5.15) can be explicitly written as $\hat{\theta}_2 = \sum_{r=1}^{n} a_r Y_{[r]r}$, where

$$a_r = \frac{\xi_r/\delta_r}{\sum_{r=1}^{n} \xi_r^2/\delta_r}, r = 1, 2, \ldots, n.$$

Remark 4.5.1. As the association parameter α in (4.5.1) is involved in the BLUE $\hat{\theta}_2$ of θ_2 and its variance, an assumption that α is known may sometimes viewed as unrealistic. Hence when α is unknown, our recommendation is to compute the sample correlation coefficient q of the observations $(X_{(r)r}, Y_{[r]r})$, $r = 1, 2, \ldots$ and consider the model (4.5.1) for a value of α equal to α_0 given by

$$\alpha_0 = \begin{cases} -1, & \text{if } q < -\frac{1}{4} \\ 4q, & \text{if } \quad -\frac{1}{4} \le q \le \frac{1}{4} \\ 1, & \text{if } q > \frac{1}{4}, \end{cases}$$

as we know that the correlation coefficient between X and Y involved in the Morgenstern type bivariate exponential random vector is equal to $\frac{\alpha}{4}$.

We have computed the the ratio $e(\hat{\theta}_2|\tilde{\theta}_2)= \frac{\mathrm{Var}(\tilde{\theta}_2)}{\mathrm{Var}(\hat{\theta}_2)}$ as the efficiency of $\hat{\theta}_2$ relative to $\tilde{\theta}_2$ for $\alpha = \pm 0.25, \pm 0.5, \pm 0.75, \pm 1.0$ and $n = 2(2)10(5)20$ and are also given in Table 4.5.1. From the table, one can easily see that $\hat{\theta}_2$ is relatively more efficient than $\tilde{\theta}_2$. Further we observe from the table that $e(\hat{\theta}_2|\tilde{\theta}_2)$ increases as n and $|\alpha|$ increase.

4.5.4. Estimation based on censored ranked-set sample

In the case of the example of pollution study on sea samples (see, Bain, 1978, P.99), sometimes if there is no tar deposit at the seashore then the correspondingly located sea sample will be censored and hence on these units the observations on Y is not measured. For ranking on X observations in a sample, the censored units are assumed to have distinct and consecutive lower ranks and the remaining units are ranked with the next higher ranks in a natural order. If in this censored scheme of ranked set sampling, k units are censored, then we may represent the ranked-set sample observations on the study variate Y as $\epsilon_1 Y_{[1]1}, \epsilon_2 Y_{[2]2}, \ldots, \epsilon_n Y_{[n]n}$ where,

$$\epsilon_i = \begin{cases} 0, & \text{if the } i^{\text{th}} \text{ unit is censored} \\ 1, & \text{otherwise} \end{cases}$$

and hence $\sum_{i=1}^n \epsilon_i = n - k$. In this case the usual ranked set sample mean is equal to $\frac{\sum_{i=1}^n \epsilon_i Y_{[i]i}}{n-k}$. It may be noted that $\epsilon_i = 0$ need not occur in a natural order for $i = 1, 2, \ldots n$. Hence if we write m_i, $i = 1, 2, \ldots n-k$ as the integers such that $1 \leq m_1 < m_2 < \cdots < m_{n-k} \leq n$ and for which $\epsilon_{m_i} = 1$, then,

$$E\left[\frac{\sum_{i=1}^n \epsilon_i Y_{[i]i}}{n-k}\right] = \theta_2 \left[1 - \frac{\alpha}{2(n+1)(n-k)} \sum_{i=1}^{n-k} (n - 2m_i + 1)\right].$$

Thus it is clear that the ranked-set sample mean in the censored case is not an unbiased estimator of the population mean θ_2. However we can construct an unbiased estimator of θ_2 based on this mean. In the following theorem we have given the constructed unbiased estimator $\theta_2^*(k)$ of θ_2 based on the ranked-set sample mean under censored situation and its variance.

Theorem 4.5.2. *Suppose that the random variable* (X, Y) *has a MTBED as defined in (4.5.1). Let* $Y_{[m_i]m_i}$, $i = 1, 2, \dots n-k$ *be the ranked-set sample observations on the study variate* Y *resulting out of censoring and ranking applied on the auxiliary variable* X. *Then an unbiased estimator of* θ_2 *based on the ranked-set sample mean* $\frac{1}{n-k}\sum_{i=1}^{n-k} Y_{[m_i]m_i}$ *is given by,*

$$\theta_2^*(k) = \frac{2(n+1)}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]}\sum_{i=1}^{n-k} Y_{[m_i]m_i}$$

and its variance is given by,

$$\mathrm{Var}[\theta_2^*(k)] = \frac{4(n+1)^2\,\theta_2^2}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k}\delta_{m_i},$$

where δ_{m_i} *is as defined in (4.5.12).*

Proof 4.5.2.

$$\begin{aligned}
E[\theta_2^*(k)] &= \frac{2(n+1)}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]}\sum_{i=1}^{n-k} E[Y_{[m_i]m_i}],\\
&= \frac{2(n+1)}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]}\sum_{i=1}^{n-k}\left[1 - \alpha\frac{n-2m_i+1}{2(n+1)}\right]\theta_2,\\
&= \frac{2(n+1)}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]}\left[n-k-\frac{\alpha}{2(n+1)}\sum_{i=1}^{n-k}(n-2m_i+1)\right]\theta_2,\\
&= \theta_2.
\end{aligned}$$

Thus $\theta_2^*(k)$ is an unbiased estimator of θ_2. The variance of $\theta_2^*(k)$ is given by,

$$\begin{aligned}
\mathrm{Var}[\theta_2^*(k)] &= \frac{4(n+1)^2}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k}\mathrm{Var}(Y_{[m_i]m_i}),\\
&= \frac{4(n+1)^2\,\theta_2^2}{\left[2(n+1)(n-k) - \alpha\sum_{i=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k}\delta_{m_i},
\end{aligned}$$

where δ_{m_i} is as defined in (4.5.12). Hence the theorem.

As a competitor of the estimator $\theta_2^*(k)$, we now propose the BLUE of θ_2 based on the censored ranked-set sample, resulting out of ranking of observations on X. Let $\mathbf{Y}_{[n]}(k) = (Y_{[m_1]m_1}, Y_{[m_2]m_2}, \ldots, Y_{[m_{n-k}]m_{n-k}})'$, then the mean vector and the dispersion matrix of $\mathbf{Y}_{[n]}(k)$ are given by,

$$E[\mathbf{Y}_{[n]}(k)] = \theta_2\, \xi(k), \tag{4.5.16}$$

$$D[\mathbf{Y}_{[n]}(k)] = \theta_2^2\, G(k), \tag{4.5.17}$$

where $\xi(k) = (\xi_{m_1}, \xi_{m_2}, \ldots, \xi_{m_{n-k}})'$, $G(k) = \operatorname{diag}(\delta_{m_1}, \delta_{m_2}, \ldots, \delta_{m_{n-k}})$.

If the parameter α involved in $\xi(k)$ and $G(k)$ is known then (4.5.16) and (4.5.17) together defines a generalized Guass-Markov set up and hence the BLUE $\hat{\theta}_2(k)$ of θ_2 is obtained as,

$$\hat{\theta}_2(k) = [(\xi(k))'(G(k))^{-1}\xi(k)]^{-1}(\xi(k))'(G(k))^{-1}\mathbf{Y}_{[n]}(k) \tag{4.5.18}$$

and

$$\operatorname{Var}(\hat{\theta}_2(k)) = [(\xi(k))'(G(k))^{-1}\xi(k)]^{-1}\theta_2^2. \tag{4.5.19}$$

On substituting the values of $\xi(k)$ and $G(k)$ in (4.5.18) and (4.5.19) and simplifying we get

$$\hat{\theta}_2(k) = \frac{\sum_{i=1}^{n-k}(\xi_{m_i}/\delta_{m_i})Y_{[m_i]m_i}}{\sum_{i=1}^{n-k}\xi_{m_i}^2/\delta_{m_i}} \tag{4.5.20}$$

and

$$\operatorname{Var}(\hat{\theta}_2(k)) = \frac{1}{\sum_{i=1}^{n-k}\xi_{m_i}^2/\delta_{m_i}}\theta_2^2.$$

Remark 4.5.2. In the expression for the BLUE's $\hat{\theta}_2$ given in (4.5.15) and $\hat{\theta}_2(k)$ given in (4.5.20) the quantities ξ_r and δ_r for $1 \le r \le n$ are all non-negative, and consequently the coefficients of ranked-set sample observations are also non-negative. Thus the BLUE's $\hat{\theta}_2$ and $\hat{\theta}_2(k)$ of θ_2 are always non-negative. Thus unlike in certain situation of BLUE's where one encounters with inadmissible estimators, the estimators given by $\hat{\theta}_2$ and $\hat{\theta}_2(k)$ using ranked set sample are admissible estimators.

Remark 4.5.3. Since both the BLUE $\hat{\theta}_2(k)$ and the unbiased estimator $\theta_2^*(k)$ based on the censored ranked-set sample utilize the distributional property of the parent distribution they lose the usual robustness property. Hence in this case the BLUE $\hat{\theta}_2(k)$ shall be considered as a more preferable estimators than $\theta_2^*(k)$.

4.5.5. Unbalanced multistage ranked-set sampling

Al-Saleh and Al-Kadiri (2000) have extended first the usual concept of RSS to double stage ranked-set sampling (DSRSS) with an objective of increasing the precision of certain estimators of the population when compared with those obtained based on usual RSS or using random sampling. Al-Saleh and Al-Omari (2002) have further extended DSRSS to multistage ranked-set sampling (MSRSS) and shown that there is increase in the precision of estimators obtained based on MSRSS when compared with those based on usual RSS and DSRSS. The MSRSS (in r stages)procedure is described as follows:

(1) Randomly select n^{r+1} sample units from the target population, where r is the number of stages of MSRSS.

(2) Allocate the n^{r+1} selected units randomly into n^{r-1}sets, each of size n^2.

(3) For each set in step (2), apply the procedure of ranked-set sampling method to obtain a (judgment) ranked-set, of size n; this step yields n^{r-1} (judgment) ranked-sets, of size n each.

(4) Arrange n^{r-1} ranked-sets of size n each, into n^{r-2} sets of n^2 units each and without doing any actual quantification, apply ranked-set sampling method on each set to yield n^{r-2} second stage ranked-sets of size n each.

(5) This process is continued, without any actual quantification, until we end up with the *rth* stage (judgment) ranked-set of size n.

(6) Finally, the n identified elements in step (5) are now quantified for the variable of interest.

Instead of judgment method of ranking at each stage if there exists an auxiliary variate on which one can make measurement very easily and exactly and if the auxiliary variate is highly correlated with the variable of interest, then we can apply ranking based on these measurements to obtain the ranked-set units at each stage of MSRSS. Then on the finally selected units one can make measurement on the variable of primary interest. In this section we deal with the MSRSS by assuming that the random variable (X, Y) has a MTBED as defined in (4.5.1), where Y is the variable of primary interest and X is an auxiliary variable. In Section 4.5, we have considered a method for estimating θ_2 using the $Y_{[r]r}$ measured on the study variate Y on the the unit having rth smallest value observed on the auxiliary variable X, of the rth sample $r = 1, 2, \ldots, n$ and hence the RSS considered there was balanced. Abo-Eleneen and Nagaraja (2002) have shown that in a bivariate sample of size n arising from MTBED the concomitant of largest order statistic possesses the maximum Fisher information on θ_2 whenever $\alpha > 0$ and the concomitant of smallest order statistic possesses the maximum Fisher information on θ_2 whenever $\alpha < 0$. Hence in this section, first we consider $\alpha > 0$ and carry out an unbalanced MSRSS with the help of measurements made on an auxiliary variate to choose the ranked-set and then estimate θ_2 involved in MTBED based on the measurement made on the variable of primary interest. At each stage and from each set we choose an unit of a sample with the largest value on the auxiliary variable as the units of ranked-sets with an objective of exploiting the maximum Fisher information on the ultimately chosen ranked-set sample.

Let $U_i^{(r)}$, $i = 1, 2, \ldots, n$ be the units chosen by the (r stage) MSRSS. Since the measurement of auxiliary variable on each unit $U_i^{(r)}$ has the largest value, we may write $Y_{[n]i}^{(r)}$ to denote the value measured on the variable of primary interest on $U_i^{(r)}$, $i = 1, 2, \ldots, n$. Then it is easy to see that each $Y_{[n]i}^{(r)}$ is the concomitant of the largest order statistic of n^r independently and identically distributed bivariate random variables with MTBED. Moreover $Y_{[n]i}^{(r)}$, $i = 1, 2, \ldots, n$ are also independently distributed with pdf given by (see, Scaria and Nair, 1999)

$$f_{[n]i}^{(r)}(y;\alpha) = \frac{1}{\theta_2} e^{-\frac{y}{\theta_2}} \left\{ 1 + \alpha \left(\frac{n^r - 1}{n^r + 1} \right) \left(1 - 2e^{-\frac{y}{\theta_2}} \right) \right\}. \tag{4.5.21}$$

Thus the mean and variance of $Y_{[n]i}^{(r)}$ for $i = 1, 2, \ldots, n$ are given below.

$$E[Y_{[n]i}^{(r)}] = \theta_2 \left[1 + \frac{\alpha}{2} \left(\frac{n^r - 1}{n^r + 1} \right) \right], \tag{4.5.22}$$

$$\mathrm{Var}[Y_{[n]i}^{(r)}] = \theta_2^2 \left[1 + \frac{\alpha}{2} \left(\frac{n^r - 1}{n^r + 1} \right) - \frac{\alpha^2}{4} \left(\frac{n^r - 1}{n^r + 1} \right) \right]. \tag{4.5.23}$$

If we denote

$$\xi_{n^r} = 1 + \frac{\alpha}{2} \left(\frac{n^r - 1}{n^r + 1} \right) \tag{4.5.24}$$

and

$$\delta_{n^r} = 1 + \frac{\alpha}{2} \left(\frac{n^r - 1}{n^r + 1} \right) - \frac{\alpha^2}{4} \left(\frac{n^r - 1}{n^r + 1} \right) \tag{4.5.25}$$

then (4.5.22) and (4.5.23) can be written as

$$E[Y_{[n]i}^{(r)}] = \theta_2 \xi_{n^r} \tag{4.5.26}$$

and

$$\text{Var}[Y_{[n]i}^{(r)}] = \theta_2^2 \delta_{n^r}. \tag{4.5.27}$$

Let $\mathbf{Y}_{[n]}^{(r)} = (Y_{[n]1}^{(r)}, Y_{[n]2}^{(r)}, \cdots, Y_{[n]n}^{(r)})'$, then by using (4.5.26) and (4.5.27) we get the mean vector and dispersion matrix of $\mathbf{Y}_{[n]}^{(r)}$ as,

$$E[\mathbf{Y}_{[n]}^{(r)}] = \theta_2 \xi_{n^r} \mathbf{1} \tag{4.5.28}$$

and

$$D[\mathbf{Y}_{[n]}^{(r)}] = \theta_2^2 \delta_{n^r} \mathbf{I}, \tag{4.5.29}$$

where $\mathbf{1}$ is the column vector of n ones and $\mathbf{I}$ is an identity matrix of order n. If $\alpha > 0$ involved in ξ_{n^r} and δ_{n^r} is known then (4.5.28) and (4.5.29) together define a generalized Gauss-Markov setup and hence the BLUE of θ_2 is obtained as

$$\hat{\theta}_2^{n(r)} = \frac{1}{n\xi_{n^r}} \sum_{i=1}^{n} Y_{[n]i}^{(r)}, \tag{4.5.30}$$

with variance given by

$$\text{Var}(\hat{\theta}_2^{n(r)}) = \frac{\delta_{n^r}}{n(\xi_{n^r})^2} \theta_2^2. \tag{4.5.31}$$

If we take $r = 1$ in the MSRSS method described above, then we get the usual single stage unbalanced RSS. Then the BLUE $\hat{\theta}_2^{n(1)}$ of θ_2 is given by

$$\hat{\theta}_2^{n(1)} = \frac{1}{n\xi_n} \sum_{i=1}^{n} Y_{[n]i},$$

with variance

$$\text{Var}(\hat{\theta}_2^{n(1)}) = \frac{\delta_n}{n(\xi_n)^2} \theta_2^2,$$

where we write $Y_{[n]r}$ instead of $Y_{[n]r}^{(1)}$ and it represent the measurement on the variable of primary interest of the unit selected in the RSS. Also ξ_n and δ_n are obtained by putting $r = 1$ in (4.5.24) and (4.5.25) respectively.

We have evaluated the ratio $e(\hat{\theta}_2^{n(1)}|\tilde{\theta}_2) = \frac{\text{Var}(\tilde{\theta}_2)}{\text{Var}(\hat{\theta}_2^{n(1)})}$ for $\alpha = 0.25(0.25)1$; $n = 2(2)10(5)20$ as a measure of efficiency of our estimator $\hat{\theta}_2^{n(1)}$ relative to the MLE $\tilde{\theta}_2$ of θ_2 based on n observations and are also provided in Table 4.5.2. From the table, one can see that the efficiency increases with increase in α and n. Moreover the efficiency of the estimator $\hat{\theta}_2^{n(1)}$ is larger than the estimator θ_2^* based on RSS mean and the BLUE $\hat{\theta}_2$ based on usual RSS.

Al-Saleh (2004) has considered the steady-state RSS by letting r to $+\infty$. If we apply the steady-state RSS to the problem considered in this paper then the asymptotic distribution of $Y_{[n]i}^{(r)}$ is given by the pdf

$$f_{[n]i}^{(\infty)}(y;\alpha) = \frac{1}{\theta_2} e^{-\frac{y}{\theta_2}} \left\{1 + \alpha\left(1 - 2e^{-\frac{y}{\theta_2}}\right)\right\}. \tag{4.5.32}$$

From the definition of our unbalanced MSRSS it follows that $Y_{[n]i}^{(\infty)}$, $i = 1, 2, \ldots, n$ are independently and identically distributed random variables each with pdf as defined in (4.5.32). Then $Y_{[n]i}^{(\infty)}$, $i = 1, 2, \ldots, n$ may be regarded as unbalanced steady-state ranked-set sample of size n. Then the mean and variance of $Y_{[n]i}^{(\infty)}$ for $i = 1, 2, \ldots, n$ are given below.

$$E[Y_{[n]i}^{(\infty)}] = \theta_2\left[1 + \frac{\alpha}{2}\right]$$

and

$$\mathrm{Var}[Y_{[n]i}^{(\infty)}] = \theta_2^2\left[1 + \frac{\alpha}{2} - \frac{\alpha^2}{4}\right].$$

Let $\mathbf{Y}_{[n]}^{(\infty)} = (Y_{[n]1}^{(\infty)}, Y_{[n]2}^{(\infty)}, \cdots, Y_{[n]n}^{(\infty)})'$. Then the BLUE $\hat{\theta}_2^{n(\infty)}$ based on $\mathbf{Y}_{[n]}^{(\infty)}$ and the variance of $\hat{\theta}_2^{n(\infty)}$ is obtained by taking the limit as $r \to \infty$ in (4.5.30) and (4.5.31) respectively and are given by

$$\hat{\theta}_2^{n(\infty)} = \frac{1}{n\left[1 + \frac{\alpha}{2}\right]} \sum_{i=1}^{n} Y_{[n]i}^{(\infty)},$$

and

$$\mathrm{Var}(\hat{\theta}_2^{n(\infty)}) = \frac{\left[1 + \frac{\alpha}{2} - \frac{\alpha^2}{4}\right]}{n\left[1 + \frac{\alpha}{2}\right]^2}\theta_2^2. \tag{4.5.33}$$

From (4.5.10) and (4.5.33) we get the efficiency of $\hat{\theta}_2^{n(\infty)}$ relative to $\tilde{\theta}_2$ by taking the ratio of Var($\tilde{\theta}_2$) with Var($\hat{\theta}_2^{n(\infty)}$) and is given by

$$e(\hat{\theta}_2^{n(\infty)}|\tilde{\theta}_2) = \frac{\mathrm{Var}(\tilde{\theta}_2)}{\mathrm{Var}(\hat{\theta}_2^{n(\infty)})},$$
$$= \frac{I_{\theta_2}^{(-1)}(\alpha)\left[1+\frac{\alpha}{2}\right]^2}{\left[1+\frac{\alpha}{2}-\frac{\alpha^2}{4}\right]}.$$

Thus the efficiency $e(\hat{\theta}_2^{n(\infty)}|\tilde{\theta}_2)$ is free of the sample size n. That is, for a fixed α, $e(\hat{\theta}_2^{n(\infty)}|\tilde{\theta}_2)$ is a constant for all n. We have evaluated the value $e(\hat{\theta}_2^{n(\infty)}|\tilde{\theta}_2)$ for $\alpha = 0.25(0.25)1$ and are presented in Table 4.5.3. From the table one can see that the efficiency of $\hat{\theta}_2^{n(\infty)}$ increases as α increases. Moreover, the estimator $\hat{\theta}_2^{n(\infty)}$ possesses the highest efficiency amoung the other estimators of θ_2 proposed in this paper and the value of the efficiencies ranges from 1.1382 to 1.7093.

As mentioned earlier, for MTBED the concomitant of smallest order statistic possesses the maximum Fisher information on θ_2 whenever $\alpha < 0$. Therefore when $\alpha < 0$ we consider an unbalanced MSRSS in which at each stage and from each set we choose an unit of a sample with the smallest value on the auxiliary variable as the units of ranked-sets with an objective of exploiting the maximum Fisher information on the ultimately chosen ranked-set sample.

Let $Y_{[1]i}^{(r)}$, $i = 1, 2, \ldots, n$ be the value measured on the variable of primary interest on the units selected at the *rth* stage of the unbalanced MSRSS. Then it is easy to see that each $Y_{[1]i}^{(r)}$ is the concomitant of the smallest order statistic of n^r , independently and identically distributed bivariate random variables with MTBED. Moreover $Y_{[1]i}^{(r)}$, $i = 1, 2, \ldots, n$ are also independently distributed with pdf given by

$$f_{[1]i}^{(r)}(y;\alpha) = \frac{1}{\theta_2}e^{-\frac{y}{\theta_2}}\left\{1-\alpha\left(\frac{n^r-1}{n^r+1}\right)\left(1-2e^{-\frac{y}{\theta_2}}\right)\right\} \tag{4.5.34}$$

Clearly from (4.5.21) and (4.5.34) we have

$$f_{[1]i}^{(r)}(y;\alpha) = f_{[n]i}^{(r)}(y;-\alpha). \tag{4.5.35}$$

and hence $E(Y_{[n]i}^{(r)})$ for $\alpha > 0$ and $E(Y_{[1]i}^{(r)})$ for $\alpha < 0$ are identically equal. Similarly $\mathrm{Var}(Y_{[n]i}^{(r)})$ for $\alpha > 0$ and $\mathrm{Var}(Y_{[1]i}^{(r)})$ for $\alpha < 0$ are identically equal. Consequently if $\hat{\theta}_2^{1(r)}$ is the BLUE of θ_2, involved in MTBED for $\alpha < 0$, based on the unbalanced MSRSS observations $Y_{[1]i}^{(r)}, i = 1, 2, \ldots, n$, then the coefficients of $Y_{[1]i}^{(r)}, i = 1, 2, \ldots, n$ in the BLUE $\hat{\theta}_2^{1(r)}$ for $\alpha < 0$ are same as the coefficients of $Y_{[n]i}^{(r)}, i = 1, 2, \ldots, n$ in the BLUE $\hat{\theta}_2^{n(r)}$ for $\alpha > 0$. Further, we have $\mathrm{Var}(\hat{\theta}_2^{1(r)}) = \mathrm{Var}(\hat{\theta}_2^{n(r)})$ and hence $\mathrm{Var}(\hat{\theta}_2^{1(1)}) = \mathrm{Var}(\hat{\theta}_2^{n(1)})$ and $\mathrm{Var}(\hat{\theta}_2^{1(\infty)}) = \mathrm{Var}(\hat{\theta}_2^{n(\infty)})$, where $\hat{\theta}_2^{1(1)}$ is the BLUE of θ_2, for $\alpha < 0$ based on the usual unbalanced single stage RSS observations $Y_{[1]i}, i = 1, 2, \ldots, n$ and $\hat{\theta}_2^{1(\infty)}$ is the BLUE of θ_2, for $\alpha < 0$ based on the unbalanced steady-state RSS observations $Y_{[1]i}^{(\infty)}, i = 1, 2, \ldots, n$. We have obtained the efficiency $e(\hat{\theta}_2^{1(1)}|\tilde{\theta}_2)$ of the BLUE $\hat{\theta}_2^{1(r)}$ relative to $\tilde{\theta}_2$, the MLE of θ_2 for $\alpha = -0.25, -0.5, -0.75, -1$; $n = 2(2)10(5)20$ and are incorporated in Table 4.5.2. Similarly as in the case of $\hat{\theta}_2^{n(\infty)}$ for a fixed α the efficiency $e(\hat{\theta}_2^{1(\infty)}|\tilde{\theta}_2)$ is same for all n. We have evaluated $e(\hat{\theta}_2^{1(\infty)}|\tilde{\theta}_2)$ for $\alpha = -0.25, -0.5, -0.75, -1$ and are incorporated in Table 4.5.3. From the table, one can see that efficiency increases as $|\alpha|$ increases and the value of efficiency ranges from 1.1382 to 1.7161.

Remark 4.5.4. If (X, Y) follows an MTBED with pdf defined in (4.5.1), then the correlation coefficient between X and Y is given by,

$$\mathrm{Corr}(X, Y) = \frac{\alpha}{4}, -1 \leq \alpha \leq 1.$$

Clearly when $|\alpha|$ goes to 1, correlation coefficient between X and Y is high. Thus, using the ranks of Y to induce the ranks of X becomes more accurate. Thus when $|\alpha|$ is large (that is α tends to ± 1) we see that the ranked-set sample obtained based on the ranking made on X becomes more informative for making inference on θ_2 than the case with small values of $|\alpha|$. From the table we notice that for a given sample size the efficiencies of all estimators increase as $|\alpha|$ increases. Consequently we note that more information on θ_2 can be extracted from the ranked-set sample when $|\alpha|$ is large subject to $|\alpha| \leq 1$. Thus we conclude that concomitant ranking is more effective in estimating θ_2 when the absolute value of the association parameter α is large (that is when α tends to ± 1).

Table 4.5.2
Efficiencies of the estimators θ_2^*, $\hat{\theta}_2$, $\hat{\theta}_2^{n(1)}$ and $\hat{\theta}_2^{1(1)}$relative to $\tilde{\theta}_2$ of θ_2 involved in Morgenstern type bivariate exponential distribution

n	α	$e(\theta_2^*\|\tilde{\theta}_2)$	$e(\hat{\theta}_2\|\tilde{\theta}_2)$	$e(\hat{\theta}_2^{n(1)}\|\tilde{\theta}_2)$	α	$e(\theta_2^*\|\tilde{\theta}_2)$	$e(\hat{\theta}_2\|\tilde{\theta}_2)$	$e(\hat{\theta}_2^{1(1)}\|\tilde{\theta}_2)$
2	0.25	0.9994	0.9994	1.0410	-0.25	0.9994	0.9994	1.0410
	0.50	0.9970	0.9970	1.0795	-0.50	0.9974	0.9974	1.0800
	0.75	0.9911	0.9911	1.1130	-0.75	0.9926	0.9926	1.1147
	1.00	0.9767	0.9768	1.1349	-1.00	0.9806	0.9807	1.1394
4	0.25	1.0008	1.0008	1.0782	-0.25	1.0008	1.0008	1.0782
	0.50	1.0026	1.0027	1.1613	-0.50	1.0030	1.0031	1.1618
	0.75	1.0038	1.0044	1.2466	-0.75	1.0054	1.0060	1.2485
	1.00	0.9996	1.0019	1.3263	-1.00	1.0036	1.0059	1.3316
6	0.25	1.0014	1.0014	1.0948	-0.25	1.0014	1.0014	1.0948
	0.50	1.0051	1.0052	1.1994	-0.50	1.0055	1.0056	1.1998
	0.75	1.0094	1.0105	1.3111	-0.75	1.0109	1.0120	1.3131
	1.00	1.0097	1.0140	1.4224	-1.00	1.0137	1.0180	1.4281
8	0.25	1.0018	1.0018	1.1042	-0.25	1.0018	1.0018	1.1042
	0.50	1.0064	1.0066	1.2213	-0.50	1.0068	1.0070	1.2218
	0.75	1.0125	1.0139	1.3490	-0.75	1.0141	1.0154	1.3511
	1.00	1.0154	1.0211	1.4800	-1.00	1.0195	1.0252	1.4860
10	0.25	1.0020	1.0020	1.1103	-0.25	1.0020	1.0020	1.1103
	0.50	1.0073	1.0075	1.2355	-0.50	1.0077	1.0079	1.2360
	0.75	1.0145	1.0161	1.3739	-0.75	1.0161	1.0176	1.3760
	1.00	1.0191	1.0259	1.5184	-1.00	1.0232	1.0300	1.5245
15	0.25	1.0023	1.0023	1.1189	-0.25	1.0023	1.0023	1.1189
	0.50	1.0085	1.0088	1.2560	-0.50	1.0089	1.0092	1.2565
	0.75	1.0173	1.0192	1.4100	-0.75	1.0189	1.0208	1.4122
	1.00	1.0243	1.0328	1.5747	-1.00	1.0284	1.0370	1.5810
20	0.25	1.0024	1.0024	1.1234	-0.25	1.0024	1.0024	1.1234
	0.50	1.0091	1.0095	1.2669	-0.50	1.0095	1.0099	1.2674
	0.75	1.0188	1.0209	1.4295	-0.75	1.0204	1.0225	1.4317
	1.00	1.0270	1.0366	1.6054	-1.00	1.0311	1.0408	1.6118

Table 4.5.3
Efficiencies of the estimators $\hat{\theta}_2^{n(\infty)}$ and $\hat{\theta}_2^{1(\infty)}$relative to $\tilde{\theta}_2$

α	$e(\hat{\theta}_2^{n(\infty)}\|\tilde{\theta}_2)$	α	$e(\hat{\theta}_2^{1(\infty)}\|\tilde{\theta}_2)$
0.25	1.1382	-0.25	1.1382
0.50	1.3028	-0.50	1.3033
0.75	1.4943	-0.75	1.4960
1.00	1.7093	-1.00	1.7161

References

Abo-Eleneen, Z.A. and Nagaraja, H.N. (2002). Fisher information in an order statistic and its concomitant. *Annals of Institute of Statistical Mathematics*, **54**, 667-680.

Al-Saleh, M.F. (2004). Steady-state ranked set sampling and parametric inference. *Journal of Statistical Planning and Inference*, **123**, 83-95.

Al-Saleh, M.F. and Al-Kadiri, M. (2000). Double ranked set sampling. *Statistics and Probability Letters*, **48**, 205-212.

Al-Saleh, M.F. and Al-Omari, A. (2002). Multistage ranked set sampling. *Journal of Statistical Planning and Inference*, **102**, 273-286.

Bain, L.J. (1978). *Statistical Analysis of Reliability and Life Testing Models: Theory and Methods*. Marcel Dekker, New York.

Barnett, V. and Moore, K. (1997). Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering. *Journal of Applied Statistics*, **24**, 697-710.

Chen, Z. (2000). The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families. *Statistica Sinica*, **10**, 247-263.

Chen, Z. and Bai, Z. (2000). The optimal ranked-set sampling scheme for parametric families. *Sankhya Series A*, **46**, 178-192.

Chen, Z., Bai, Z. and Sinha, B. K. (2004). *Lecture Notes in Statistics, Ranked-Set Sampling, Theory and Applications*. Springer-Verlag, New York.

David, H.A. and Nagaraja, H.N (2003). *Order Statistics*, Wiley, New York.
Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Distributions in Statistics: Continuous Multivariate Distributions,* Wiley, New York.

Lam, K., Sinha, B.K. and Wu, Z. (1994). Estimation of a two-parameter exponential distribution using ranked-set sample. *Annals of the Institute of Statistical Mathematics*, **46**, 723-736.

Lam, K., Sinha, B.K. and Wu, Z. (1996). Estimation of location and scale parameters of a logistic distribution using ranked-set sample. In H.N. Nagaraja, P. K. Sen and D.F. Morrison(Ed.) *Statistical Theory and Applications: Papers in Honor of Herbert A. David*. New-York: Springer.

McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked-sets. *Australian Journal of Agricultural Research*, **3**, 385-390.

Modarres, R. and Zheng, G. (2004). Maximum likelihood estimation of dependence parameter using ranked-set sampling. *Statistics and Probability Letters*, **68**, 315-323.

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

Stokes, S. L. (1977). ranked-set sampling with concomitant variables. *Communications in Statistics; Theory and Methods*, **6**, 1207-1211.

Stokes, S. L. (1980). Inference on the correlation coefficient in bivariate normal populations from ranked-set Samples. *Journal of the American Statistical Association*, **75**, 989-995.

Stokes, S. L. (1995). Parametric ranked-set sampling. *Annals of the Institute of Statistical Mathematics*, **47**, 465-482.

Zheng, G. and Modarres, R. (2006). A robust estimate of correlation coefficient for bivariate normal distribution using ranked-set sampling. *Journal of Statistical Planning and Inference*, **136**, 298-309.