ISSN: 2160-5912 Vol. 5, No. 4, July 2013

Scientific Research

Intelligent Information Management

ISSN: 2160-5912

www.scirp.org/journal/iim

Journal Editorial Board

ISSN: 2160-5912 (Print), 2160-5920 (Online)
http://www.scirp.org/journal/iim

Intelligent Information Management, **2013, 5, 103-139**
Published Online July 2013 in SciRes (http://www.scirp.org/journal/iim/)

TABLE OF CONTENTS

Volume 5 Number 4 July 2013

Intelligent Information Management (IIM)

Journal Information

SUBSCRIPTIONS

The *Intelligent Information Management* (Online at Scientific Research Publishing, www.SciRP.org) is published bimonthly by Scientific Research Publishing, Inc., USA.

Subscription rates:

Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: iim@scirp.org

Intelligent Information Management, **2013, 5, 103-116**
doi:10.4236/iim.2013.54011 Published Online July 2013 (http://www.scirp.org/journal/iim)

A Video Game Based on Optimal Control and Elementary Statistics

Marco Giacinti[1], Francesca Mariani[2], Maria Cristina Recchioni[2], Francesco Zirilli[2]
[1]Dipartimento di Scienze Sociali "D. Serrani", Università Politecnica delle Marche, Ancona, Italy
[2]CERI-Centro di Ricerca Previsione Prevenzione e Controllo dei Rischi Geologici,
Università di Roma "La Sapienza", Roma, Italy
Email: m.giacinti@univpm.it, fra_mariani@libero.it, m.c.recchioni@univpm.it, zirilli@mat.uniroma1.it

Received April 15, 2013; revised May 16, 2013; accepted June 10, 2013

ABSTRACT

The video game presented in this paper is a prey-predator game where two preys (human players) must avoid three predators (automated players) and must reach a location in the game field (the computer screen) called preys' home. The game is a sequence of matches and the human players (preys) must cooperate in order to achieve the best performance against their opponents (predators). The goal of the predators is to capture the preys, which are the predators try to have a "rendez vous" with the preys, using a small amount of the "resources" available to them. The score of the game is assigned following a set of rules to the prey team, not to the individual prey. In some situations the rules imply that to achieve the best score it is convenient for the prey team to sacrifice one of his components. The video game pursues two main purposes. The first one is to show how the closed loop solution of an optimal control problem and elementary statistics can be used to generate (game) actors whose movements satisfy the laws of classical mechanics and whose behaviour simulates a simple form of intelligence. The second one is "educational", in fact the human players in order to be successful in the game must understand the restrictions to their movements posed by the laws of classical mechanics and must cooperate between themselves. The video game has been developed having in mind as players for children aged between five and thirteen years. These children playing the video game acquire an intuitive understanding of the basic laws of classical mechanics (Newton's dynamical principle) and enjoy cooperating with their teammate. The video game has been experimented on a sample of a few dozen children. The children aged between five and eight years find the game amusing and after playing a few matches develop an intuitive understanding of the laws of classical mechanics. They are able to cooperate in making fruitful decisions based on the positions of the preys (themselves), of the predators (their opponents) and on the physical limitations to the movements of the game actors. The interest in the game decreases when the age of the players increases. The game is too simple to interest a teenager. The game engine consists in the solution of an assignment problem, in the closed loop solution of an optimal control problem and in the adaptive choice of some parameters. At the beginning of each match, and when necessary during a match, an assignment problem is solved, that is the game engine chooses how to assign to the predators the preys to chase. The resulting assignment implies some cooperation among the predators and defines the optimal control problem used to compute the strategies of the predators during the match that follows. These strategies are determined as the closed loop solution of the optimal control problem considered and can be thought as a (first) form of artificial intelligence (AI) of the predators. In the optimal control problem the preys and the predators are represented as point masses moving according to Newton's dynamical principle under the action of friction forces and of active forces. The equations of motion of these point masses are the constraints of the control problem and are expressed through differential equations. The formulation of the decision process through optimal control and Newton's dynamical principle allows us to develop a game where the effectiveness and the goals of the automated players can be changed during the game in an intuitive way simply modifying the values of some parameters (*i.e.* mass, friction coefficient, $\cdots$). In a sequence of game matches the predators (automated players) have "personalities" that try to simulate human behaviour. The predator personalities are determined making an elementary statistical analysis of the points scored by the preys in the game matches played and consist in the adaptive choice of the value of a parameter (the mass) that appears in the differential equations that define the movements of the predators. The values taken by this parameter determine the behaviour of the predators and their effectiveness in chasing the preys. The predators personalities are a (second) form of AI based on elementary statistics that goes beyond the intelligence used to chase the preys in a match. In a sequence of matches the predators using this second form of AI adapt their behaviour to the preys' behaviour. The video game can be downloaded from the website: http://www.ceri.uniroma1.it/ceri/zirilli/w10/.

Keywords: Video Game; Differential Games; Mechanical Dynamical System; Closed Loop Optimal Control

1. Introduction

This paper shows how optimal control models and elementary statistics can be used to simulate intelligent behaviour in a video game involving two teams of actors. The movements of the actors are restricted by the physical laws of classical mechanics. The video game is a prey-predator game where two human players (preys) face the opposition of three automated players (predators). The preys pursue the goal of avoiding the predators and of reaching a location on the game field called preys' home. The predators pursue the goal of capturing the preys (*i.e.* of having a "rendez vous" with the preys) using a small amount of the "resources" available to them.

The game is a sequence of matches and the human players (preys) must cooperate in order to achieve the best score against their opponents (predators). The score of the game is the sum of the scores of the matches. The score of a match is assigned following a set of rules described in Subsection 4.3 to the prey team, not to the individual players. The situation where, at the end of a match, a prey is captured and the other one has reached the preys' home generates a score higher than the score generated by the situation where, at the end of the match, the two preys have avoided the rendez vous with the predators but neither of them has reached the preys' home. This fact implies that in some circumstances to achieve the best score for the prey team it is convenient to sacrifice one of the preys. This is an incentive to cooperate for the human players. It is believed that this kind of cooperation has several positive effects on the psychological attitude of the players especially if they are aggressive young children [1].

The video game is addressed to human players aged between five and thirteen years and can be used to stimulate children to develop an intuitive understanding of the physical laws involved in the game. The children aged between five and thirteen years are usually not aware of the relations among mass, friction forces, active forces and trajectories implied by Newton's dynamical principle. Playing with the video game and observing the behaviour of its actors these children develop an intuitive understanding of the dynamical laws. This understanding allows them to make effective decisions about the strategy to implement to reach the goals of the game. During a game match some simple forms of cooperation, communication and strategic reasoning are employed by the preys and by the predators. These features make the video game amusing.

The video game can be downloaded from the website: http://www.ceri.uniroma1.it/ceri/zirilli/w10/.

1.1. The Advantages of Using Optimal Control and Elementary Statistics to Develop a Video Game

The main advantage of using optimal control in the development of a video game is that allows to manage the personalities (*i.e.* the behaviours) of the automated players (sometimes called "autonomous synthetic characters") in a simple way and allows to define trajectories satisfying physical laws. The effectiveness of the predators in chasing the preys depends on the choice of the utility functional of the optimal control problem considered in the video game (see Formula (7)) and on the equations of motion that are the constraints of the control problem. In particular it depends on the choice of the parameters that appear in the utility functional and in the equations of motion. The equations of motion express the physical laws involved in the game. That is modeling the video game via an optimal control problem generates predators that are able to chase the preys and whose movements depend on their physical properties (*i.e.* mass, friction coefficient, $\cdots$) and on the trajectories of the preys. Note that there is no preassigned scheme for the movement of the predators.

The automated players (autonomous synthetic characters) developed in this video game have two forms of artificial intelligence (AI). The first one is used to chase the preys during a game match. The second one is used when a sequence of matches is played to simulate in the autonomous characters the behaviour in similar circumstances of human players. This second AI form is articulated in three different levels. The first one guarantees that the predators behave in a slightly different way in each match that is they behave depending on the humor of the moment. The second level consists in the dynamic change of the game difficulty to avoid that the players are discouraged or bored as a consequence of the fact that the game is too difficult or too easy. The last level simulates the emotional reaction of the predators corresponding to the human satisfaction or anger for a sequence of victories or of defeats (see Subsection 4.8 for technical details).

The mathematical model that determines the movement of the predators during a match, that is that determines the first form of AI, is the optimal control problem. In the optimal control problem the game actors are represented as point masses. The action of the actors (preys and predators) during a game match can be seen as an attempt to solve a "rendez vous" problem for a set of point masses that satisfy Newton's dynamical principle. These point masses are subject to external forces and to friction forces. The movements of the preys and of the predators are obtained as solution of differential equations that depend on the external forces. These differential equations express Newton's dynamical principle. The external forces acting on the preys are chosen by the human players using joypads and those acting on the predators are computed by the game engine as the closed loop solution of the optimal control problem considered. The utility functional of the control problem depends on the prey positions and its minimization corresponds to pursuing the goal of chasing the preys using a small quantity of the resources available to the predators, which are using a small quantity of the external forces acting on the predators. The utility functional chosen determines a form of cooperation among the predators. In a sequence of matches the autonomous characters of the video game (predators) modify their behaviour depending on the behaviour of the human players (preys) that confront them, which is depending on the points scored by the human players in the matches played. This second form of AI is built using elementary statistics and exploiting the properties of the control problem employed in the video game. In fact the statistical analysis of the points scored by the human players in a sequence of matches is used to adapt the game in relation to the human players actually playing. This approach has something in common with the idea introduced in [2] of defining a difficulty evaluation function for a video game. In the video game presented in this paper thank to the mathematical model (*i.e.* the optimal control problem) used it is easy to establish a relation between the ability of the human players and the values of the parameters of the model. That is, it is easy to modify dynamically the game difficulty level.

1.2. Research Background

In recent years the interest in autonomous synthetic characters has been constantly growing due to the increased importance of computer animation and of interactive media. In the technical literature several approaches have been suggested to build this kind of characters see, for example, Prada and Paiva [3] and Millington [4].

In video games these characters are associated to automated players. In [4] the behaviour of the automated players is based on pathfinding algorithms (Section 4.2 of [4]), and the decision making capability of these players is based on the analysis of random trees. Moreover in [4] Chapter 7 the question of how to manage learning experiences of human and automated players based on AI and on elementary statistics is discussed (see Section 7.3.2 of [4]).

Indeed one of the most relevant challenges in video game AI is to create automated players having an appropriate level of effectiveness in pursuing their goals. This is often done implementing a finite state machine (see Section 5.3 of [4]), that is presenting the game at several "levels" of difficulty. This approach suffers from some drawbacks, such as the fact that it is not possible to adjust strategies or difficulty dynamically during the game. Recently an alternative approach to the finite state machine has been introduced (see [5] and the references therein). This approach is called Dynamic Difficulty Adjustment (DDA) and consists in changing automatically in real time parameters, scenarios or behaviours in the video game based on player's ability and performance. In [5] in a simple prey-predator game the DDA approach is implemented through a Monte-Carlo tree search.

However the problem of defining a measure of the game difficulty is in general an open question and several approaches have been suggested. One of them consists in introducing a difficulty evaluation function defined as a conditional probability (see [2] and the reference therein) on the "lose" or "win" events. In [2] the authors underline that, once established a measure of the game difficulty, in general it is not easy to change the game difficulty dynamically using simple mathematical expressions of the game parameters. The use of optimal control and of elementary statistics to drive the difficulty adjustment algorithm proposed in this paper is an example of a practical tool to handle this issue. In fact using these mathematical models we are able to generate video games with a level of difficulty calibrated on the individual players. In the following Sections and in [6] it is explained why modeling a video game through an optimal control problem or a differential game is a way to parameterize the difficulty of the video game in an intuitive way.

Another important issue when a multitude of automated players is considered is the simulation of group behaviour. The simulation of group behaviour of a multitude of characters is inspired by the study of human behaviour in the social sciences and usually involves some elementary mathematical models limited to path-finding algorithms and to some elementary formulae that define the movements of the synthetic characters in the game scene (see [3,7-10]). The problem of building mathematical models to describe the behaviour of a multitude of autonomous characters interacting as a group has also been studied outside the context of AI and of social sciences. For example it has been studied in natural sciences, in fact in 1987 Reynolds [11] developed a mathematical model of the flocking behaviour of a group of flying creatures (birds or insects). Since then several mathematical models of the social behaviour of several species of animals, such as birds, fishes, insects, have been studied. These models are based on dynamical systems and usually do not involve optimal control or differential games. The three automated players of the video game presented here have some cooperative behaviour (see Section 1.3) but are not a multitude and do not exhibit a group behaviour.

The movements of the actors in prey-predator video games are realized most of the times using mathematical approaches simpler than the optimal control approach used here (see [9,10,12]). For example in [9,10] the movements of automated players are based on ordinary differential equations and on several heuristic rules that define the movements of articulated bodies consisting of rigid links connected by joints. In absence of joints each link has six degrees of freedom, three translational and three rotational. In [9,10] the authors model the movements of these bodies, but do not deal with the problem of generating players able to make decisions and to cooperate to pursue a goal. In [12] the problems of modeling movements and of simulating a decision making process are dealt together and the results presented have some similarities to the ones presented in this paper. In fact in [12] each actor of the game is characterized by its position vector and by a unit vector called "heading direction". The heading directions are not necessarily the velocities of the actors in the game scene and their dynamics is described by a system of ordinary differential equations that depends on the interactions among actors. These interactions are modeled through a function inspired by an analogy with potential theory (see [12] for further details). Similarly in [13] the authors deal with the problem of modeling prey-predator interactions using mathematical models that go beyond simple attraction/repulsion models. In fact in [13] a mathematical model obtained modifying the Lotka-Volterra two species equations [14] is proposed. The Lotka-Volterra two species equations describe the interaction between two species in an ecosystem, a predator and a prey species, and consists in two ordinary differential equations, the first one describes how the prey population changes and the second one describes how the predator population changes. In [13] the authors propose an algorithm to reproduce collective crowd behaviour based on a Lotka-Volterra two species model and on some steering rules that try to simulate realistic crowd behaviours. Recall that being a member of a crowd changes the behaviour of the individual. The models used in [12,13] suffer from the drawback of not being able to change the game difficulty dynamically.

Finally video games where basic physics principles hold are presented in [15,16]. In [15,16] the authors are motivated by robotics. Their work is based on the ALERT (Active Learning Environments with Robotic Tangibles) systems. The ALERT systems have been developed for e-learning purposes and recently have been used to develop educational video games [16]. However these are special cases and the majority of the commercial games violate basic physics principles and give to the players the feeling that physics principles can be disregarded. For example often in video games movement is considered only from the kinematic point of view and dynamic principles are neglected. Note that the models proposed in [12,13,15,16] do not involve optimal control problems or differential games.

The video game proposed in this paper is an extension of the video game presented in [6]. In fact both these games are based on mathematical models involving differential equations and define actors whose movements satisfy Newton's dynamical principle. In [6] the strategies implemented by the predators are suboptimal solutions of the control problem that models the prey-predator dynamics. Moreover in [6] the video game does not adapt the predators effectiveness to their human opponents. The video game presented in [6] is implemented at three different levels of difficulty. The levels of difficulty considered are given. The predators of [6] have only the intelligence needed to chase the prey in a game match complying with Newton's dynamical principle. The video game presented in this paper uses optimal solutions of the control problem and has a mechanism to adjust its difficulty to the ability of the human players.

Let us conclude this subsection noting that in mathematics rendez vous problems involving two actors or two sets of actors whose dynamics is defined by systems of differential equations have been widely studied. The most common mathematical models used to study these problems are optimal control models or differential games (see, for example, Athans [17], Isaacs [18]). Differential games are the natural model to describe situations where two sets of actors having conflicting goals moving on trajectories implicitly defined by differential equations must choose their strategies respectively to maximize or to minimize a cost functional. The game developed here is an attempt to investigate the potentiality of these mathematical models in context of video games.

1.3. Interactive and Cooperative Behaviours in the Video Game

Let us summarize the prey-predator interaction and cooperation that will be explained in detail in Section 4. Let us group the game actors in two teams, which are the team 1 actors, the preys, and the team 2 actors, the predators.

Let us explain more in detail of the interaction and cooperation among the game actors. We have already mentioned that the cooperation between the human players is induced by the rules used to assign the game score (see Subsection 4.3 for further details) and in particular by the fact that the score is given to the human team not to the individual players.

At the beginning of the first match played the video game engine opens a window on the computer screen for each human player (see **Figure 1**). During the game execution each prey is always at the center of its own window. Each window shows a neighborhood of the prey at its center. Moreover in each window there is a second view of the game scene that shows the game scene on a global scale, which is on a scale that shows the location of all the actors of the game and the location of the preys' home (see **Figure 1**). Also this global view of the game scene has the prey in its center and in the sequel is attributed to the "radar" of the player.

Each human player before beginning to play a match observes the initial scene of the game on his radar and chooses a strategy to play the match that follows cooperating with the other human player. For example, observing the predators' positions each human player decides if it is convenient for his survival and for the score of the human team in the following match to move towards or outwards the preys' home.

When the game match starts each human player watches the time evolution of the game scene in his window and uses a joypad to implement the strategy chosen. In each match the human players (preys) and the automated players (predators) have a finite amount of time to achieve their goals. After the expiration of this time the match ends. At the end of each match depending on the outcome of the match the human players score some points (see Subsection 4.3). Under some conditions a new match can be played, that is the game can be resumed from a new initial scene. During the match the interaction between preys and predators takes place.

In the actual implementation of the game the rendez vous of a predator with a prey during a game match corresponds to the fact that the distance between the prey and the predator becomes smaller than a given tolerance. Similarly during the game match a prey reaches its home when the distance of the prey from its home becomes smaller than a given tolerance. The preys' home is represented as a brown disk in the game field (see **Figure 1**). A prey after having a rendez vous with a predator or after reaching the preys home is removed from the game scene for the remaining part of the match.

During the execution of a game match the data generated by the human players using the joypads are transferred continuously in time to the game engine. The game engine processes these data in real time, that is in real time computes the new positions and velocities of the preys, and computes the closed loop solution of the optimal control problem considered to determine the new

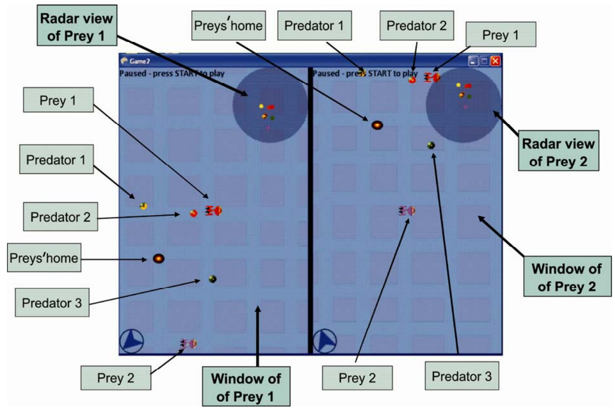

Figure 1. An initial scene of the game.

positions and velocities of the predators. Finally the game engine redraws preys and predators positions in the windows of the players and updates the global views of the game scene shown by the radars of the players. Moreover at the beginning of each match and when necessary during a match the game engine solves the assignment problem.

During a sequence of matches, depending on the statistical analysis of the points scored by the human players in the matches played, some parameters that define the behaviour of the predators are changed. In the implementation of the video game presented in Section 4 the parameters changed are the masses of the predators, however it is possible to change several other parameters appearing in the model defining predators with more sophisticated behavioural patterns than those implemented in Section 4. This mechanism makes the game difficulty adaptive and simulate an elementary form of personality in the predators.

1.4. Outline of the Paper

In Section 2 we discuss the "rendez vous" problem and its formulation as an optimal control problem. In Section 3 we present the closed loop solution of the optimal control problem formulated in Section 2. In Section 4 we explain the video game, in particular we explain the cooperative behaviour of the preys and of the predators and we show how the personalities of the predators are generated and adaptively changed during a sequence of matches (see Section 4.8). Moreover we present some technical and practical details of the implementation of the video game. In Section 5 we summarize the experience of a few dozen of children that have played with the video game and we draw some conclusions.

2. The Mathematical Model Used in the Game Engine

Let us formulate the optimal control problem whose solution is implemented in the engine of the video game. The position of the five actors of the video game (two preys and three predators) is given by five two-dimensional vectors in the Euclidean plane. These vectors and all the other vectors appearing in this paper are column vectors. The preys and the predators are point masses moving in the Euclidean plane. We use the vectors $\underline{x}^i$, $i=1,2,$ to denote the prey positions and the vectors $\underline{y}^j$, $j=1,2,3,$ to denote the predator positions. Let $T>0$ be a constant, we assume that the game matches begin at time $t=0$ and end at time $t=T>0$. That is the game matches take place for t such that $0\le t\le T$. The vectors $\underline{x}^i=\underline{x}^i(t)$, $0\le t\le T$, $i=1,2,$ $\underline{y}^j=\underline{y}^j(t),$ $0\le t\le T,$ $j=1,2,3,$ describe the movements of the game actors in the Euclidean plane during a match, that is they describe the action going on in the game scene during a match.

The human players animate the preys located in $\underline{x}^i$, $i=1,2,$ and in the time interval $0\le t\le T$ pursue the goals of avoiding the rendez vous with the predators located in $\underline{y}^j$, $j=1,2,3,$ and of reaching as soon as possible the preys home (that is a fixed location in the plane). These goals are pursued acting on two joypads. The action on the joypads of the human players defines two vector functions $\underline{u}^i=\underline{u}^i(t),$ $0\le t\le T,$ $i=1,2,$ taking values in the Euclidean plane. For $0\le t\le T$ the vector $\underline{u}^i=\underline{u}^i(t)$ is the active force acting on the point mass representing the prey i, $i=1,2$.

Let $0<\tau\le T,$ given the forces $\underline{u}^i=\underline{u}^i(t),$ $0\le t\le\tau$, $i=1,2$, the game engine computes the corresponding movements of the preys $\underline{x}^i=\underline{x}^i(t)$, $0\le t\le\tau$, $i=1,2$. In the video game for every t such that $0\le t\le T$, at time t the positions of the preys up to time t, that is $\underline{x}^i=\underline{x}^i(\tau_1)$, $0\le\tau_1\le t$, $i=1,2,$ are known to the automated players. The preys have a similar knowledge of the preys and predators positions.

The automated players (predators) located in $\underline{y}^j=\underline{y}^j(t)$, $0\le t\le T$, $j=1,2,3,$ in the time interval $0\le t\le T$ pursue the goals of having a rendez vous with the preys located in $\underline{x}^i=\underline{x}^i(t)$, $0\le t\le T$, $i=1,2,$ using a "small quantity" of the active forces at their disposal. These active forces are denoted with $\underline{v}^j=\underline{v}^j(t)$, $0\le t\le T,$ $j=1,2,3,$ and are functions of t taking values in the Euclidean plane. The force $\underline{v}^j=\underline{v}^j(t)$, $0\le t\le T,$ acts on the point mass representing the predator $j,$ $j=1,2,3.$

2.1. Newton's Dynamical Principle Applied to Preys and Predators Movement

The functions $\underline{x}^i(t)$, $0\le t\le T$, $i=1,2,$ are solutions of the following differential equations:

$$m_{x^i}\ddot{\underline{x}}^i=\alpha_{x^i}\dot{\underline{x}}^i+\underline{u}^i,\ 0<t<T, i=1,2, \tag{1}$$

with the initial conditions:

$$\underline{x}^i(0)=\underline{x}^{0,i},\ i=1,2, \tag{2}$$

$$\dot{\underline{x}}^i(0)=\underline{v}_x^{0,i},\ i=1,2, \tag{3}$$

where $\dot{*}$, $\ddot{*}$ denote respectively the first and second order derivatives with respect to t of $*$ and m_{x^i}, α_{x^i}, $i=1,2$, are real constants. The constants $m_{x^i}>0$, $i=1,2$, are the masses of the preys. The constants $\alpha_{x^i}<0$, $i=1,2,$ are the coefficients of the friction forces acting on the preys. The vectors $\underline{x}^{0,i}$, $\underline{v}_x^{0,i}$ are given and represent respectively the initial position and the initial velocity of the prey i, $i=1,2$. Note that the forces $\underline{u}^i=\underline{u}^i(t)$, $0<t<T,$ $i=1,2$, appear in (1) as forcing terms.

The predators functions $\underline{y}^j(t)$, $0\le t\le T$, $j=1,2,3,$ are solutions of the following differential equations:

$$m_{y^j}\ddot{\underline{y}}^j=\alpha_{y^j}\dot{\underline{y}}^j+\underline{v}^j,\ 0<t<T, j=1,2,3, \tag{4}$$

with the initial conditions:

$$\underline{y}^j(0)=\underline{y}^{0,j},\ j=1,2,3, \tag{5}$$

$$\dot{\underline{y}}^j(0)=\underline{v}_y^{0,j},\ j=1,2,3, \tag{6}$$

where $m_{y^j}>0$, $j=1,2,3,$ are the masses of the predators and $\alpha_{y^j}<0$, $j=1,2,3,$ are the coefficients of the friction forces acting on the predators. The vectors $\underline{y}_y^{0,j}$, $\underline{v}_y^{0,j}$ are given and represent respectively the initial position and the initial velocity of the predator j, $j=1,2,3.$ Note that the forces $\underline{v}^j=\underline{v}^j(t)$, $0<t<T$, $j=1,2,3,$ appear in (4) as forcing terms and that they are determined as solution of the optimal control problem formulated below.

The differential Equations (1) and (4) are interpreted as Newton's dynamical principle $(m\underline{a}=\underline{F})$ for two sets of point masses representing respectively the preys and the predators. In this sense they are inspired by classical mechanics and the movements implicitly defined by them, that is the trajectories of the preys and of the predators, satisfy basic physics laws. Let us point out that for every τ such that $0<\tau\le T$ once known the data $\underline{u}^i(t)$, $0\le t<\tau$, $i=1,2$, it is possible to derive an explicit formula that gives the corresponding trajectories of the preys $\underline{x}^i(t)$, $0\le t<\tau$, $i=1,2$, up to time $t=\tau$. This formula reduces the solution of the initial value problem (1), (2), (3), that defines $\underline{x}^i(t),$ $0\le t\le T$, to the evaluation of an integral that contains the control function $\underline{u}^i(t)$, $0\le t<\tau\le T$, $i=1,2$. A similar statement holds for the initial value problems (4)-(6).

2.2. The Optimal Control Problem and the Dynamics of the Game Actors

Given $\underline{x}^i(t)$, $0\le t\le T$, $i=1,2,$ we determine $\underline{v}^j(t)$,

$0 \le t \le T$, $j = 1,2,3,$ as solution of the optimal control problem:

$$\min_{\underline{v}^j, j=1,2,3} \left\{ \sum_{j=1}^{3} \frac{\mu_{v^j}}{2} \int_0^T \left\langle \underline{v}^j(t), \underline{v}^j(t) \right\rangle \mathrm{d}t + \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{\mu_{x^i,y^j}}{2} \int_0^T \left\langle \underline{x}^i(t) - \underline{y}^j(t), \underline{x}^i(t) - \underline{y}^j(t) \right\rangle \mathrm{d}t \right\}, \tag{7}$$

subject to the constraints (4), (5), (6). In (7) $\langle \cdot,\cdot \rangle$ denotes the scalar product in the Euclidean plane. Later $\langle \cdot,\cdot \rangle$ will denote the scalar product in a generic Euclidean space not necessarily two dimensional. The quantities μ_{x^i,y^j}, $i = 1,2$, $j = 1,2,3,$ and μ_{v^j}, $j = 1,2,3,$ that appear in (7) are nonnegative constants satisfying some conditions that will be specified later. Note that in order to be usable in the video game engine the solution of the control problem (7), (4), (5), (6) must be in closed loop form. That is for any τ such that $0 \le \tau \le T$, from the knowledge of $\underline{x}^i(t)$, $0 \le t < \tau$, $i = 1,2$, we must determine $\underline{v}^j(t)$, $0 \le t < \tau$.

Let us explain why the initial value problem (1), (2), (3) and the control problem (7), (4), (5), (6) are a legitimate mathematical model of the game described in the Introduction. In fact the control functions $\underline{u}^i(t)$, $0 \le t \le T$, $i = 1,2,$ are chosen by the human players acting on the joypads based on the observation of the scene on the computer screen and are interpreted as forces acting on the preys. These forces define implicitly (through the initial value problems (1), (2), (3)) the strategy followed by the human players to achieve their goals. The choice of the control functions $\underline{u}^i(t)$, $0 \le t \le T$, $i = 1,2$, is reflected in the control problem (7), (4), (5), (6) through the presence in (7) of the functions $\underline{x}^i(t)$, $0 \le t \le T$, $i = 1,2.$ This implies that the choice of the control functions made by the human players determines the behaviour of the automated players obtained solving the control problem (7), (4), (5), (6). The control functions $\underline{v}^j(t)$, $0 \le t \le T$, $j = 1,2,3,$ solution of (7), (4), (5), (6) define implicitly through the initial value problems (4), (5), (6) the movements of the predators. Let us explain more in detail the cost functional (7). For $i = 1,2$, $j = 1,2,3$ when the nonnegative constant μ_{x^i,y^j} is greater than zero making small the term $\frac{\mu_{x^i,y^j}}{2} \int_0^T \left\langle \underline{x}^i(t) - \underline{y}^j(t), \underline{x}^i(t) - \underline{y}^j(t) \right\rangle \mathrm{d}t$ of (7) corresponds to trying to realize the rendez vous between the predator j and the prey i and similarly when the nonnegative constant μ_{v^j} is greater than zero making small the term $\frac{\mu_{v^j}}{2} \int_0^T \left\langle \underline{v}^j(t), \underline{v}^j(t) \right\rangle \mathrm{d}t$ of (7) corresponds to the attempt of using a "small quantity" of the control function $\underline{v}^j$ (*i.e.* the available resource) during the rendez vous. In the game implementation we choose $\mu_{v^j} = 1$, $j = 1,2,3$. The choice of the constants μ_{x^i,y^j}, $i = 1,2,$ $j = 1,2,3,$ corresponds to the solution of the assignment problem mentioned previously, that is corresponds to the assignment of the preys to the predators. In fact for $i = 1,2,$ $j = 1,2,3$ when $\mu_{x^i,y^j} > 0$ the prey i is assigned to the predator j (that is the predator j chases the prey i), when $\mu_{x^i,y^j} = 0$ the prey i is not assigned to the predator j. The value of the constant μ_{x^i,y^j} is a measure of the "attraction" of the predator j towards the prey i, $i = 1,2,$ $j = 1,2,3.$ That is for $i = 1,2,$ $j = 1,2,3$ increasing the value of the constant μ_{x^i,y^j} increases the attraction of the predator j toward the prey i. In the implementation of the game when $\mu_{x^i,y^j} > 0$ we choose $\mu_{x^i,y^j} = 2$, $i = 1,2,$ $j = 1,2,3.$ The assignment of the preys to the predators done by the game engine at the beginning of each match satisfies the following rules: 1) each predator must chase only one prey, 2) each prey must be chased by at least one predator. These rules imply a kind of cooperation between the predators. The assignment of the preys to the predators is redone during a match when a prey is captured by a predator (*i.e.* there is a rendez vous) or when a prey reaches the preys' home. In these cases only one prey remains active in the game scene and all the predators chase it. Note that when a rendez vous takes place or a prey reaches the prey home the differential models (1), (2), (3) and (7), (4), (5), (6) are reformulated in an obvious way to take care of the situation determined by these events.

The match begins at time $t = 0$ and ends at time $t = t_1$, $0 < t_1 < T$, where t_1 is the smallest time value such that for $t = t_1$ two preys have reached the preys' home or one prey has been captured and the other one has reached the prey home or two preys have been captured. In all the remaining cases the match ends when the time assigned to the match is expired, that is when $t = T$. At the beginning of each match when a new (initial) scene is proposed the game engine does the assignment of the preys to the predators (*i.e.* chooses the constants μ_{x^i,y^j}, $i = 1,2,$ $j = 1,2,3$) simply associating to each prey the nearest predator. Moreover the engine verifies that the assignment obtained in this way satisfies the rules stated above. For example when using the "nearest predator rule" two preys are assigned to the same predator the engine changes this assignment, assigning to this predator only the nearest prey and giving the other prey to the nearest predator among the remaining ones. In the assignment problem when necessary some simple choices are done. For example, degenerate cases such as the case when there are two preys at the same distance from the nearest predator are solved making random choices. The assignment chosen by the game engine satisfies always the rules stated above.

3. The "Closed Loop" Solution of the Optimal Control Problem Used in the Game Engine

Let p be a positive integer, $\mathbb{R}^p$ be the p dimensional real Euclidean space and $\underline{0}_2 \in \mathbb{R}^2$ be the two dimensional null vector. Let us solve in closed loop form the optimal control problem (7), (4), (5), (6). We define the vector of the state variables of problem (7), (4), (5), (6) at time t, $\underline{\xi}(t) \in \mathbb{R}^{12}$, $0 \le t \le T$, and the vector $\underline{\xi}^{i,a}(t) \in \mathbb{R}^{12}$, $0 \le t \le T$, that contains the position of the prey i at time t, $i = 1,2$, that is:

$$\underline{\xi}(t) = \begin{pmatrix} \underline{y}^1(t) \\ \underline{\dot{y}}^1(t) \\ \underline{y}^2(t) \\ \underline{\dot{y}}^2(t) \\ \underline{y}^3(t) \\ \underline{\dot{y}}^3(t) \end{pmatrix}, \underline{\xi}^{i,a}(t) = \begin{pmatrix} \underline{x}^i(t) \\ \underline{0}_2 \\ \underline{x}^i(t) \\ \underline{0}_2 \\ \underline{x}^i(t) \\ \underline{0}_2 \end{pmatrix}, \; 0 \le t \le T, i = 1,2. \quad (8)$$

Moreover let us introduce the vector of the control functions at time t, $\underline{w}(t) \in \mathbb{R}^6$, $0 \le t \le T$, that is:

$$\underline{w}(t) = \begin{pmatrix} \underline{v}^1(t) \\ \underline{v}^2(t) \\ \underline{v}^3(t) \end{pmatrix}, \; 0 < t < T. \quad (9)$$

Recall that the vectors defined in (8), (9) are column vectors. For later convenience we define I_2 to be the 2×2 identity matrix and O_2 to be the 2×2 null matrix. The optimal control problem (7), (4), (5), (6) can be rewritten as follows:

$$\min_{\underline{w}} \left\{ \frac{1}{2} \int_0^T \left\langle \underline{w}(t), R\underline{w}(t) \right\rangle \mathrm{d}t + \frac{1}{2} \sum_{i=1}^{2} \int_0^T \left\langle \underline{\xi}(t) - \underline{\xi}^{i,a}(t), Q_i \left(\underline{\xi}(t) - \underline{\xi}^{i,a}(t) \right) \right\rangle \mathrm{d}t \right\}, \quad (10)$$

subject to the constraints:

$$\underline{\dot{\xi}}(t) = A\underline{\xi}(t) + B\underline{w}(t), \; 0 < t < T, \quad (11)$$

with the initial condition:

$$\underline{\xi}(0) = \underline{\xi}^0 = \begin{pmatrix} \underline{y}^{1,0} \\ \underline{v}_y^{1,0} \\ \underline{y}^{2,0} \\ \underline{v}_y^{2,0} \\ \underline{y}^{3,0} \\ \underline{v}_y^{3,0} \end{pmatrix}, \quad (12)$$

where $\underline{\xi}^0 \in \mathbb{R}^{12}$ is a column vector. The matrices Q_i, $i = 1,2$, and A are 12×12 real matrices, B is a 12×6 real matrix and R is a 6×6 real matrix. These matrices are given by:

$$Q_i = \begin{pmatrix} \mu_{x^i,y^1} I_2 & O_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & \mu_{x^i,y^2} I_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & O_2 & \mu_{x^i,y^3} I_2 & O_2 \\ O_2 & O_2 & O_2 & O_2 & O_2 & O_2 \end{pmatrix}, \quad (13)$$

$i = 1,2$,

$$A = \begin{pmatrix} O_2 & I_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & \frac{\alpha_{y^1}}{m_{y^1}} I_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & O_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & I_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 & \frac{\alpha_{y^2}}{m_{y^2}} I_2 & O_2 & I_2 \\ O_2 & O_2 & O_2 & O_2 & O_2 & \frac{\alpha_{y^3}}{m_{y^3}} I_2 \end{pmatrix}, \quad (14)$$

$$B = \begin{pmatrix} O_2 & O_2 & O_2 \\ \frac{1}{m_{y^1}} I_2 & O_2 & O_2 \\ O_2 & O_2 & O_2 \\ O_2 & \frac{1}{m_{y^2}} I_2 & O_2 \\ O_2 & O_2 & O_2 \\ O_2 & O_2 & \frac{1}{m_{y^3}} I_2 \end{pmatrix}, \quad (15)$$

$$R = \begin{pmatrix} \mu_{v^1} I_2 & O_2 & O_2 \\ O_2 & \mu_{v^2} I_2 & O_2 \\ O_2 & O_2 & \mu_{v^3} I_2 \end{pmatrix}. \quad (16)$$

The matrix vector products appearing in (10), (11) are the usual rows by columns matrix, vector products.

Note that for $i = 1,2$ the matrix Q_i defines the assignment of the prey i to the predators, and that the functions $\underline{\xi}^{i,a}(t)$, $0 \le t \le T$, $i = 1,2$, are data in the optimal control problem (10), (11), (12).

It is easy to see that the first and the second term of the cost functional (10) correspond respectively to the first and the second term of the cost functional (7).

The special form of the optimal control problem (10), (11), (12) (*i.e.*: quadratic cost functional and linear constraints) guarantees that the Hamilton Jacobi equation associated to it can be reduced to a system of ordinary differential equations that can be integrated numerically (see [19] pp. 138-142) to obtain an approximate closed loop solution of the optimal control problem (10), (11), (12) or equivalently of the problem (7), (4), (5), (6). Proceeding in this way we approximate the strategy of the predators $\underline{w}^*(t)$, $0 \le t \le T$, solution of problem (10), (11), (12), as follows:

$$\underline{w}^*(t) = -R^{-1}B^{\mathrm{T}}\sum_{i=1}^{2}\left[K_{1,i}(t)\left(\underline{\xi}^*(t) - \underline{\xi}^{i,a}(t)\right) + \underline{k}_{2,i}(t)\right], \quad 0 \le t \le T, \tag{17}$$

where the optimal trajectory of the state variables $\underline{\xi}^*(t)$, $0 \le t < T$, is the solution of the following system of differential equations:

$$\begin{aligned}\dot{\underline{\xi}}^*(t) &= A\underline{\xi}^*(t) \\ &-BR^{-1}B^{\mathrm{T}}\cdot\left[\sum_{i=1}^{2}\left(K_{1,i}(t)\left(\underline{\xi}^*(t)-\underline{\xi}^{i,a}(t)\right)+\underline{k}_{2,i}(t)\right)\right], \quad 0 \le t < T,\end{aligned} \tag{18}$$

with the initial condition:

$$\underline{\xi}^*(0) = \underline{\xi}^0, \tag{19}$$

and $K_{1,i}(t)$ is a 12×12 real matrix, $\underline{k}_{2,i}(t) \in \mathbb{R}^{12}$ is a vector, $0 \le t \le T$, $i = 1,2$. The quantities $K_{1,i}(t)$, $\underline{k}_{2,i}(t)$, $0 < t < T$, $i = 1,2$, that appear in (17), (18) are solutions of the following systems of differential equations:

$$\begin{aligned}\dot{K}_{1,i}(t) &= -K_{1,i}(t)A - A^{\mathrm{T}}K_{1,i}(t) - Q_i \\ &+K_{1,i}(t)BR^{-1}B^{\mathrm{T}}K_{1,i}(t), \ 0 < t < T, i = 1,2,\end{aligned} \tag{20}$$

$$\begin{aligned}\dot{\underline{k}}_{2,i}(t) &= -A^{\mathrm{T}}\underline{k}_{2,i}(t) - K_{1,i}(t)A\underline{\xi}^{i,a}(t) \\ &+K_{1,i}(t)\dot{\underline{\xi}}^{i,a}(t), 0 < t < T, i = 1,2,\end{aligned} \tag{21}$$

with the final conditions:

$$K_{1,i}(T) = O, \ i = 1,2, \tag{22}$$

$$\underline{k}_{2,i}(T) = \underline{0}, \ i = 1,2, \tag{23}$$

where in (22), (23) O and $\underline{0}$ are respectively the 12×12 null matrix and the 12 dimensional null vector. The system of ordinary differential Equations (20) is Riccati system of differential equations. Note that the matrices $K_{1,i} = K_{1,i}(t)$, $0 < t \le T$, $i = 1,2$, solutions of (20), (22), can be computed (numerically) at time $t = 0$ integrating backward in time (20) starting from the final condition (22). This can be done, for example, using Euler method. This is possible since Equations (20), (22) do not depend on the trajectories $\underline{\xi}^{i,a}(t)$, $0 \le t \le T$, $i = 1,2$.

The vectors $\underline{k}_{2,i} = k_{2,i}(t)$, $0 < t \le T$, $i = 1,2,$ solutions of (21), (23) cannot be computed at time $t = 0$ in the same way since to do that we should know at time $t = 0$ the functions $\underline{\xi}^{i,a}(t)$, $i = 1,2,$ in the interval $0 \le t \le T$. However it is not possible to assume that at time $t = 0$ the prey trajectories contained in the functions $\underline{\xi}^{i,a}(t)$, $0 \le t \le T$, $i = 1,2$, are known in the time interval $0 \le t \le T$.

Hence in order to compute the solution of the optimal control problem (10), (11), (12) we proceed as follows:

Step 0. Decompose the interval $[0,T]$ in q subintervals, that is let $[0,T] = \bigcup_{j=0}^{q-1}[t_j, t_{j+1}]$, where $t_j = jT/q$, $j = 0,1,\cdots,q-1$; For $j = 0,1,\cdots,q-1$ do the following steps:

Step 1. Solve numerically using Euler method backward in time the system of ordinary differential Equations (20) in the interval $[t_j, t_{j+1}]$ with the null condition at $t = t_{j+1}$ and store the approximation $\tilde{K}_{1,i}(t_j)$ of $K_{1,i}(t_j)$, $i = 1,2$, obtained in this way.

Step 2. From the knowledge of $\underline{\xi}^{i,a}(t_j)$, $\dot{\underline{\xi}}^{i,a}(t_j)$, $\ddot{\underline{\xi}}^{i,a}(t_j)$ approximate $\underline{\xi}^{i,a}(t_{j+1})$ and $\dot{\underline{\xi}}^{i,a}(t_{j+1})$ respectively with $\tilde{\underline{\xi}}^{i,a}(t_{j+1})$, $\dot{\tilde{\underline{\xi}}}^{i,a}(t_{j+1})$ given by the following formulae:

$$\tilde{\underline{\xi}}^{i,a}(t_{j+1}) = \underline{\xi}^{i,a}(t_j) + \left(\frac{T}{q}\right)\dot{\underline{\xi}}^{i,a}(t_j) + \frac{1}{2}\left(\frac{T}{q}\right)^2\ddot{\underline{\xi}}^{i,a}(t_j), \tag{24}$$

$$\dot{\tilde{\underline{\xi}}}^{i,a}(t_{j+1}) = \dot{\underline{\xi}}^{i,a}(t_j) + \left(\frac{T}{q}\right)\ddot{\underline{\xi}}^{i,a}(t_j). \tag{25}$$

Solve numerically using Euler method backward in time the system of differential Equations (21) in the interval $[t_j, t_{j+1}]$, with the null condition at $t = t_{j+1}$, that is compute:

$$\begin{aligned}\tilde{\underline{k}}_{2,i}(t_j) &= \tilde{\underline{k}}_{2,i}(t_{j+1}) - (T/q)\Big(-A^{\mathrm{T}}\tilde{\underline{k}}_{2,i}(t_{j+1}) \\ &-\tilde{K}_{1,i}(t_{j+1})A\tilde{\underline{\xi}}^{i,a}(t_{j+1}) + \tilde{K}_{1,i}(t_{j+1})\dot{\tilde{\underline{\xi}}}^{i,a}(t_{j+1})\Big), i = 1,2,\end{aligned} \tag{26}$$

with the final conditions:

$$\tilde{\underline{k}}_{2,i}(t_{j+1}) = \underline{0}, \ i = 1,2, \tag{27}$$

where $\tilde{\underline{k}}_{2,i}(t_j)$ is the quantity that approximates $\underline{k}_{2,i}(t_j)$, $i = 1,2$.

Step 3. From the knowledge of an approximation $\tilde{\underline{\xi}}^*(t_j)$ of $\underline{\xi}^*(t_j)$ compute numerically an approximation $\tilde{\underline{\xi}}^*(t_{j+1})$ of $\underline{\xi}^*(t_{j+1})$ at time $t = t_{j+1}$ solving (18), (19) using Euler method forward in time, which is compute:

$$\underline{\tilde{\xi}}^*(t_{j+1}) = \underline{\tilde{\xi}}^*(t_j) + (t_{j+1} - t_j) A \underline{\tilde{\xi}}^*(t_j) - (t_{j+1} - t_j) \cdot BR^{-1} B^{\mathrm{T}} \sum_{i=1}^{3} \left[\tilde{K}_{1,i}(t_j) \left(\underline{\tilde{\xi}}^*(t_j) - \underline{\xi}^{i,a}(t_j) \right) + \underline{\tilde{k}}_{2,i}(t_j) \right]. \quad (28)$$

Recall that the initial condition at $t = t_0 = 0$ of $\underline{\xi}^*(t)$, $0 \le t \le T$, is given by:

$$\underline{\tilde{\xi}}^*(t_0) = \underline{\xi}^*(t_0) = \underline{\xi}^*(0) = \underline{\xi}^0. \quad (29)$$

Finally compute the approximation $\underline{\tilde{w}}^*(t_{j+1})$ of the optimal control $\underline{w}^*(t_{j+1})$ at time $t = t_{j+1}$ using formula (17).

Note that using the Steps 0-3 in order to approximate the optimal strategy $\underline{w}^*(t_{j+1})$ and the optimal trajectory $\underline{\xi}^*(t_{j+1})$ at time $t = t_{j+1}$ it is necessary to know only $\underline{\tilde{\xi}}^*(t_j)$ and $\underline{\tilde{\xi}}^{i,a}(t_j)$, $i = 1,2$, $j = 0,1,\cdots,q-1$.

In the implementation of the video game described in Section 4 Steps 0-3 are used with $T = 30$ seconds and $q = 300$.

4. The Video Game Implementation and the Personalities of the Predators

The video game described in the previous Sections and further specified in this Section has been implemented in C# (C sharp) language and can be downloaded from the website: http://www.ceri.uniroma1.it/ceri/zirilli/w10/. The website contains also some auxiliary material that helps the understanding of this paper.

4.1. System Requirements

The video game works under Windows XP, Vista and Windows 7 both on 32-bit and on 64 bit thank to a setup that recognizes the hardware configuration (32 or 64 bit). To play it is necessary to use the joypad Xbox 360 controller for Windows (see **Figure 2**). Each human player must have a joypad.

Figure 2. The joypad Xbox 360.

4.2. Game Setting

The human players must avoid the rendez vous with the three automated players (the predators) and must reach a given location in the plane called preys' home. The preys' home is represented by a brown disk with a yellow center on the computer screen. The game field is the entire plane and each prey has at its disposal a window on the computer screen (see **Figure 1**). During the game each prey is always at the center of its window. That is the window scrolls up, down, left and right according to the movement of the prey. In every window in the right top corner there is a global view of the game scene provided by the "radar" of the prey. This global view of the scene has the prey at its center and shows all the actors of the game and the preys' home. Finally in the left bottom corner of every window there is an arrow that indicates the direction (within the window) of the vector that joins the center of the window with the preys' home. This arrow is particularly useful when the preys' home does not appear in the local view of the game scene shown in the window to indicate to the prey the direction where to go to find its home. **Figure 1** shows the two windows of the preys as they appear on the computer screen.

The possible outcomes of a game match are: 1) two preys reach the preys' home, 2) one prey reaches the preys' home and the other prey has not been captured at time $t = T$, 3) one prey reaches the preys' home and the other prey has been captured, 4) two preys have not reached the preys' home and have not been captured at time $t = T$, *v*) one prey has not reached the preys' home and has not been captured at time $t = T$ and the other prey has been captured, 6) both preys have been captured. To these different outcomes are attributed points following the rules specified in Section 4.3.

The game consists in a sequence of matches, each match has a duration of thirty seconds, that is $T = 30$ seconds. Each match starts from its own initial scene. An initial scene is made of the location of the preys' home and of the preys and predators positions at time $t = 0$. The velocities of the preys and of the predators at time $t = 0$ are chosen to be zero. The initial scenes of the matches are generated randomly. The way of specifying the personalities of the predators in a sequence of matches is discussed in Section 4.8. At the beginning of each match the human players observe the initial scene of the game looking at their windows on the screen, in particular looking at the image of the game scene shown by their radars. Before the beginning or during a pause of a match the game action is stopped and the human players can agree on an eventually cooperative strategy to play the match. The choice of the strategy is dictated by the desire of making in the match the best point score that is compatible with the game scene and with the game rules. The strategy chosen must be implemented by the human players using the joypads in the action that takes place after starting or resuming the match.

As already said the match is over when $t = T$ or before $t = T$ when one of the following conditions are satisfied: the two preys have been captured, the two preys have reached their home, one prey has been captured and the other has reached the preys' home. The game ends when one of the human players presses the BACK button of the joypad or when the maximum number of matches allowed has been played. In the downloadable version of the game the maximum number of matches allowed is 100.

4.3. Score Rules

The score of the game is computed summing the scores made in the matches. The score of a match is computed using the following rules:

1) One point for each prey that has not reached the preys' home and has not been captured at the end of the match, that is at $t = T$,

2) Three points for each prey that has reached the preys' home during the match.

The points scored are attributed to the team of the human players not to the players individually.

Note that the score rules, the initial scene of the match, the rules used in the assignment of the preys to the predators and the personalities of the predators (see Section 4.8) determine the strategy that the human players must choose at the beginning of each match.

4.4. Radar View of the Game

The global view of the game scene shown by the radar is a disk on the right top corner of the window associated to each prey (**Figure 1**). In this view the game scene is shown completely, that is in this view of the scene the preys' home (brown disk with a yellow center), the predators (green, orange and yellow disks) and the preys (red and violet disks) are shown. Each prey is at the center of the scene shown by its radar. The view of the scene shown by the radar can be removed from the window pressing the GREEN button of the joypad.

4.5. Preys' Home

The preys' home is a location in the plane denoted with a brown disk with a yellow center (**Figure 1**). This location is generated randomly at the beginning of each match and remains unchanged during the match. The preys' home is always visible in the global view of the game scene shown by the radars and the arrow in the left bottom corner of the window of each prey indicates the direction within the window where to go to find the preys' home (**Figure 1**).

4.6. Preys Movements and Cooperation

The preys are point masses represented as colored disks (red or violet) in the global views of the scene shown on the computer screen and their movements are implicitly defined by the equations of motion (1) with the initial conditions (2), (3). In its own window the prey is the colored (red, violet) guy at the center of the window (see **Figure 1**). The masses and the friction coefficients of the preys are assigned at the beginning of the game by the game engine and they remain unchanged until the end of the game.

Five buttons of the joypad are used by the human players during the game, that is: START button, BACK button, GREEN button, top left washer and bottom right washer (see **Figure 2**).

The forces $\underline{u}^i = \underline{u}^i(t)$, $0 \le t \le T$, $i = 1, 2$, appearing in Equation (1) are chosen by the human players acting on the joypads as follows. The movement of the top left washer to the right provides a force that pushes the prey to the right of the computer screen, the movement of the top left washer up provides a force that pushes the prey upwards on the computer screen and so on. The force magnitude is directly proportional to the washer slope. That is the signals transmitted by the top left washers of the joypads to the game engine are interpreted as two pairs of real numbers q_1^i, q_2^i belonging to the range $[-1, 1]$ that can be identified respectively with the Cartesian components (in the "natural" frame of reference) of the vectors $\underline{u}^i$ appearing in Equation (1), $i = 1, 2$. These pairs of numbers are functions of time. The movements impressed on the top left washers as a function of time define the forces that the human players provide to the corresponding preys as a function of time.

Before starting a match (or before resuming a paused match) the human players must observe the game scene on their radar views and depending on the scene observed they must choose a strategy to play the match that follows. For example, if one prey is close to the preys' home but has two predators in its immediate vicinity, these predators are presumably assigned to its capture. In this case it could be convenient in order to maximize the score made in the match to move this prey far from the preys' home forcing the two predators assigned to its capture to follow it and as a consequence to move far from the preys' home. This makes easier for the other prey to reach home. The choice of the previous strategy is motivated by the fact that the score made when one prey reaches home and the other is captured or remains alive until the end of the match is higher than the score obtained when the two preys remain alive until the end of the match but neither of them reaches the preys' home. Note that in the choice of their strategy the human players must keep in mind that at the beginning of the match one of the preys will be chased by one predator and the other one will be chased by two predators. Moreover the strategy chosen must consider the assignment rule, that is must consider the fact that basically the predators are assigned to chase the prey nearest to them in the initial scene. Hence at the beginning of a match looking at the initial scene shown by the radars the human players can image which one of the preys will be chased by two predators. Moreover the human players must consider the fact that in the initial scene of a match the velocities of the preys and of the predators are zero and that this is not true in general in the scene of a paused match. After a few matches, the human players should be able to evaluate the differences between the predators. This information is relevant in the choice of the prey strategies in particular it is relevant when the predators are located near the preys' home. However as will be seen in Section 4.8 the personalities of the predators change during a sequence of matches and as a consequence the human players in the choice of their strategy must adapt their criteria to the changes of the predator personalities.

4.7. Predators Movements and Cooperation

The predators are point masses represented as colored disks (green, orange and red) on the computer screen and their movements are implicitly defined by the equations of motion (4) with the initial conditions (5), (6). Each predator pursues the goal of chasing the prey assigned to it by the game engine using a "small quantity" of the resources available to it (*i.e.* of its control function). The predators are different one from the others. That is the parameters that appear in the predator equations are different for the red, orange and green predator. As explained in Section 4.8 in the game actually implemented during a sequence of matches the masses of the predators change depending on the behaviour of the preys while the values of the other parameters have a standard value that remains unchanged.

4.8. The Predator Personalities

When a sequence of matches is played in order to simulate a predators behaviour similar to the (supposed) human behaviour in the similar circumstances an elementary statistical analysis of the points scored by the human players in the matches played is used. This mechanism provides a form of artificial intelligence to the predators that goes beyond the artificial intelligence used to chase the preys in a match. In particular out of this mechanism three specific forms of artificial intelligence (in short AI forms) are implemented. The first AI form consists in having the predators to behave in a slightly different way in each match. This corresponds to the fact that the human behaviour in similar circumstances is not always the same but, for example, depends on the humor of the moment. The second AI form consists in calibrating the effectiveness of the predators in chasing the preys according to the ability of the preys. This is done to avoid to discourage or to bore the human players as a consequence of the fact that the predators are too effective or too ineffective in chasing the preys. The third AI form wants to simulate in the predator behaviour the emotional reaction of a human player to a sequence of defeats or of victories.

Let us go into the details of how these three AI forms are implemented.

For $j=1,2,3$ we begin assigning to the mass m_{y^j} of the predator j three possible values $m_{y^j} = m^k_{y^j}$, $k=1,2,3$, such that $0 < m^1_{y^j} < m^2_{y^j} < m^3_{y^j}$. The parameters μ_{x^i,y^j}, $i=1,2$, $j=1,2,3$, are chosen solving the assignment problem and take the values zero or two. The parameters α_{y^j}, μ_{v^j}, $j=1,2,3$, are assigned at the beginning of the game and kept constant during the matches. We have $\mu_{v^j}=1$, $j=1,2,3$, and $\alpha_{y^1}=-0.4$ (pink predator), $\alpha_{y^2}=-0.7$ (yellow predator) and $\alpha_{y^3}=-1$ (green predator).

To implement the first AI form we do the following: at the beginning of each match a random variable h_j is sampled, $j=1,2,3$. The random variable h_j assigns to the mass of the predator j the values $m^k_{y^j}$, $k=1,3$, with probability $1/4$ and the value $m^2_{y^j}$ with probability $1/2$, $j=1,2,3$.

To implement the second AI form, once every ten matches, the game engine checks the score made by the human players in the last ten matches and compares it with the maximum score attainable in the matches. If the score made is smaller than 20% of the maximum score attainable the masses of the predators are increased of 20%, that is $m^k_{y^j} = m^k_{y^j}(1+0.2)$, $k=1,2,3$, $j=1,2,3$, if the score made is greater than 70% of the maximum score attainable the masses of the predators are decreased of 20%, that is $m^k_{y^j} = m^k_{y^j}(1-0.2)$, $k=1,2,3$, $j=1,2,3$. Let us point out that increasing the mass of a predator makes it slower in its movement and, as a consequence, makes easier for the preys to avoid the "rendez vous" with this predator. Similarly decreasing the mass of a predator makes it faster in its movement and, as a consequence, makes more difficult for the preys to avoid the "rendez vous" with this predator. That is increasing the masses of the predators makes the game easier for the human players and decreasing the masses of the predators makes the game more difficult for the human players. In this way the difficulty of the game increases or decreases depending on the ability of the human players. Moreover the random behaviour of the masses of the predators that vary between the values m^k_j, $k=1,2,3$, depending on the value sampled from the random variable h_j, $j = 1,2,3$, makes impossible to know precisely the ability of the predators when planning a strategy at the beginning of a match.

Let us call defeat a match where two preys have been captured and victory a match where both preys have reached the preys' home. To realize the third AI form we check the number n_d of consecutive defeats and the number n_v of consecutive victories of the human players. When $n_v = 5$ the masses of the predators are decreased of 30%. This simulates the fact that the predators are "hungry" due to the repeated victories of the human players and become aggressive. When $n_d = 5$ the masses of all predators are increased of 30%. This simulates the fact that the predators are satisfied due to the repeated defeats of the human players and become relaxed.

5. Experimental Results and Conclusions

We have done an experiment on a sample of a few dozens of young players aged between five and thirteen years to see the reaction of these players to the exposure to the video game. The previous experiences with video games of these children were very different. The children were divided in three groups. The first group was made of children who had never played video games, the second one was made of children having a little experience with video games and the third one was made of children used to play video games. The players of the first group were five-six years old, the age of the players of the remaining groups was between five and thirteen years old. The experiment has shown that children aged between five and eight years belonging to the first two groups enjoy playing the video game and in the first matches played find difficult to cooperate. However, after some matches, these young players understand the presence of restrictions to their movements and to the movements of their opponents coming from Newton's dynamical principle and develop strategies coherent with these restrictions. An interesting finding is that even the youngest players (five-six years old) after playing a few matches develop an intuitive understanding of Newton's dynamical principle and consequently are able to improve their ability to escape from the pursuit with the predators and to cooperate with their teammate. Playing the video game is an enjoyable experience for the first two groups of children. In fact the simplicity and the dynamic change of the level of difficulty of the game encourage the players of the first group and the cooperation amuses the players of the second group. The players of the third group, in particular the teenager players, after a while get bored by the video game. In fact these players are already acquainted with the richness of commercial video games and are disappointed by the simplicity of the video game presented here.

Concluding the video game at least for the younger and inexperienced players is a practical tool to understand intuitively some basic physical laws and to develop a cooperative attitude with their teammate.

Moreover the video game is an example of how mathematical models taken from optimal control and elementary statistics can be used to develop prey-predator games where the actors' movements satisfy Newton's dynamical principle and the automated actors simulate a form of intelligent behaviour that changes dynamically depending on the behaviour of the human players during the game. No "*a priori*" schemes are used to determine the automated players' behaviour.

REFERENCES

[1] A. K. Bay-Hinitz, R. F. Peterson and H. R. Quilitch, "Cooperative Games: A Way to Modify Aggressive and Cooperative Behaviours in Young Children," *Journal of Applied Behavior Analysis*, Vol. 27, No. 3, 1994, pp. 435-446. doi:10.1901/jaba.1994.27-435

[2] M. V. Aponte, G. Levieux and S. Natkin, "Scaling the Level of Difficulty in Single Player Video Games," *Proceedings of the* 8*th International Conference on Entertainment Computing*, Paris, 3-5 September 2009, S. Natkin and J. Dupire, Eds., ICEC 5709, 2009, pp. 24-35.

[3] R. Prada and A. Paiva, "Teaming up Humans with Autonomous Synthetic Characters," *Artificial Intelligence*, Vol. 173, No. 1, 2009, pp. 80-103. doi:10.1016/j.artint.2008.08.006

[4] I. Millington, "Artificial Intelligence for Games," Elsevier, Morgan Kaufmann Publications, 2006.

[5] L. Sha, S. He, J. Wang, J. Yang, Y. Gao, Y. Zhang and X. Yu, "Creating Appropriate Challenge Level Game Opponent by the Use of Dynamic Difficulty Adjustment," 2010 6*th International Conference on Natural Computation* (*ICNC* 2010), *IEEE Circuits and Systems Society*, 2010, pp. 3897-3901.

[6] M. Giacinti, F. Mariani, M. C. Recchioni and F. Zirilli, "A Video Game Based on Elementary Equations," *Annals of Mathematics and Artificial Intelligence*, 2012.

[7] N. Beume, H. Danielsiek, T. Hein, B. Naujoks, N. Piatkowski, R. Stüer, A. Thom and S. Wessing, "Towards Intelligent Team Composition and Maneuvering in Real-Time Strategy Games," *IEEE Transactions on Computational Intelligence and AI in Games*, Vol. 2, No. 2, 2010, pp. 82-98. doi:10.1109/TCIAIG.2010.2047645

[8] D. B. Fogel, T. J. Haysa and D. R. Johnson, "A Platform for Evolving Intelligently Interactive Adversaries," *Biosystems*, Vol. 85, No. 1, 2006, pp. 72-83. doi:10.1016/j.biosystems.2006.02.010

[9] V. Kokkevis, "Practical Physics for Articulated Characters," *Game Developers Conference* 2004, San Jose, 2004, pp. 1-16. http://www.red3d.com/cwr/games/#ai-papers

[10] I. Millington, "Game Physics Engine Developments," Elsevier Inc., San Francisco, 2007.

[11] C. W. Reynolds, "Flocks, Herds, and Schools: A Distribu-

ted Behavioral Model," *Computer Graphics*, Vol. 21, No. 4, 1987, pp. 25-34. doi:10.1145/37402.37406

[12] S. I. Nishimura and T. Ikegami, "Emergence of Collective Strategies in a Prey-Predator Game Model," *Artificial Life*, Vol. 3, No. 4, 1997, pp. 243-260. doi:10.1162/artl.1997.3.4.243

[13] S. Kim, C. Hoffman and J. M. Lee, "An Experimental in Rule-Based Crowd Behaviour for Intelligent Games," *ICCIT*, 2009 *IEEE Fourth International Conference on Computer Sciences and Convergence Information Technologies*, Seoul, 24-26 November 2009, pp. 410-415. doi:10.1109/ICCIT.2009.239

[14] K. Gopalsamy, "Exchange of Equilibria in Two Species Lotka-Volterra Competition Models," *The Journal of the Australian Mathematical Society* (*Series B*), Vol. 24, No. 2, 1982, pp. 160-170. doi:10.1017/S0334270000003659

[15] B. Lahey, W. Burleson, C. N. Jensen, N. Freed, P. Lu and K. Muldner, "Integrating Video Games and Robotic Play in Physical Environments," *Proceedings of the* 2008 *ACM SIGGRAPH Symposium on Video Games*, Los Angeles, ACM Publishing, New York, 2008, Vol. 1, pp. 153-170.

[16] B. Lahey, N. Freed, P. Lu, C. N. Jensen, K. Muldner and W. Burleson, "Human-Robot Interactions to Promote Play and Learning," *Proceedings of the* 8*th International Conference on Interaction Design and Children* (*IDC* 2009), Como, ACM Publishing, New York, 2009, pp. 280-281.

[17] M. Athans, "On Optimal Allocation and Guidance Laws for Linear Interception and Rendez Vous Problems," *IEEE Transactions on Aerospace and Electronics Systems*, Vol. AES7, No. 5, 1971, pp. 843-853. doi:10.1109/TAES.1971.310324

[18] R. Isaacs, "Differential Games," Dover Publication, New York, 1999.

[19] T. L. Friesz, "Dynamic Optimization and Differential Games," *International Series in Operations Research & Management Science*, Vol. 135, Springer, New York, 2010, pp. 138-142.

Intelligent Information Management, **2013, 5, 117-125**
doi:10.4236/iim.2013.54012 Published Online July 2013 (http://www.scirp.org/journal/iim)

Estimation Based on Progressive First-Failure Censored Sampling with Binomial Removals*

Ahmed A. Soliman[1], Ahmed H. Abd Ellah[2], Nasser A. Abou-Elheggag[2], Rashad M. El-Sagheer[3]
[1]Faculty of Science, Islamic University, Madinah, Saudi Arabia
[2]Mathematics Department, Sohag University, Sohag, Egypt
[3]Mathematics Department, Faculty of Science, Al-Azhar University, Nasr-City, Cairo, Egypt
Email: Rashadmath@yahoo.com

Received April 27, 2013; revised May 28, 2013; accepted June 15, 2013

ABSTRACT

In this paper, the inference for the Burr-X model under progressively first-failure censoring scheme is discussed. Based on this new censoring were the number of units removed at each failure time has a discrete binomial distribution. The maximum likelihood, Bootstrap and Bayes estimates for the Burr-X distribution are obtained. The Bayes estimators are obtained using both the symmetric and asymmetric loss functions. Approximate confidence interval and highest posterior density interval (HPDI) are discussed. A numerical example is provided to illustrate the proposed estimation methods developed here. The maximum likelihood and the different Bayes estimates are compared via a Monte Carlo simulation study.

Keywords: Burr-X Distribution; Progressive First-Failure Censored; Bayesian and Non-Bayesian Estimations; Loss Function; Bootstrap; Random Removals

1. Introduction

Censoring is common in life-distribution work because of time limits and other restrictions on data collection. Censoring occurs when exact lifetimes are known only for a portion of the individuals or units under study, while for the remainder of the lifetimes information on them is partial. However, when the lifetimes of products are very high, the experimental time of a type II censoring life test can be still too long. A generalization of type II censoring is progressive type II censoring, which is useful when the loss of live test units at points other than the termination point is unavoidable. Johnson [1] described a life test in which the experimenter might decide how to group the test units into several sets, each as an assembly of test units, and then run all the test units simultaneously until occurrence of the first failure in each group. Such a censoring scheme is called first-failure censoring. Wu and Kuş [2] obtained maximum likelihood estimates, exact confidence intervals and exact confidence regions for the parameters of Weibull distribution under the progressive first-failure censored sampling. Note that a first-failure censoring scheme is terminated when the first failure in each set is observed. If an experimenter desires to remove some sets of test units before observing the first failures in these sets this life test plan is called a progressive first-failure censoring scheme which recently was introduced by Wu and Kuş [2]. Recently, the estimation of Parameters from different lifetime distribution based on progressive type II censored samples is studied by several authors including Gupta *et al.* [3], Childs and Balakrishnan [4], Siu keung tse *et al*. [5], Mosa and Jaheen [6], Ng *et al.* [7], Wu and Chang [8], Balakrishnan *et al.* [9], Wu [10], Soliman [11], and Sarhan and Abuammoh [12]. But in some reliability experiments, the number of patients dropped out the experiment cannot be pre-fixed and it is random. In such situations, the progressive censoring schemes with random removals are needed. Therefore, the purpose of this paper is to develop a Bayes estimation (symmetric and asymmetric loss functions) for the parameters of Burr-X distribution under the progressive first-failure censoring plan with random removals and construct the bootstrap confidence interval for the parameters.

If X follows a Burr-X distribution, then the probability density function (pdf) and cumulative distribution

*Mathematics Subject Classification: 6_2No_5; 6_2F_{10}.

function (cdf) of X are given respectively by

$$f(x) = 2\theta x \exp(-x^2)\left(1-\exp(-x^2)\right)^{\theta-1}, \quad x,\theta>0, \quad (1)$$

$$F(x) = \left(1-\exp(-x^2)\right)^{\theta}, \quad x,\theta>0. \quad (2)$$

The rest of this paper is organized as follows. In Section 2, we describe the formulation of a progressive first-failure censoring scheme as described by Wu and Kuş [2]. The point estimation of the parameters of Burr-X distribution and binomial distribution based on the progressive first-failure censoring scheme is investigated in Section 3. In Section 4, we discuss the approximate interval estimation and highest posterior density interval (HPDI) for the Burr-X distribution under the progressive first-failure censored sampling plan. A numerical examples are presented in Section 5, for illustration. In Section 6 we provide some simulation results in order to give an assessment of the performance of the estimation method.

2. A Progressive First-Failure Censoring Scheme

In this section, first-failure censoring is combined with progressive censoring as in Wu and Kuş [2]. Suppose that n independent groups with k items within each group are put in a life test, R_1 groups and the group in which the first failure is observed are randomly removed from the test as soon as the first failure (say $X^{\mathbf{R}}_{1:m:n:k}$) has occurred, R_2 groups and the group in which the second failure is observed are randomly removed from the test as soon as the second failure (say $X^{\mathbf{R}}_{2:m:n:k}$) has occurred, and finally $R_m\ (m \le n)$ groups and the group in which the $m-th$ failure is observed are randomly removed from the test as soon as the $m-th$ failure (say $X^{\mathbf{R}}_{m:m:n:k}$) has occurred. The $X^{\mathbf{R}}_{1:m:n:k} < X^{\mathbf{R}}_{2:m:n:k} < \cdots < X^{\mathbf{R}}_{m:m:n:k}$ are called progressively first-failure-censored order statistics with the progressive censoring scheme $\mathbf{R}$. It is clear that $n = m + R_1 + R_2 + \cdots + R_m$. If the failure times of the $n \times k$ items originally in the test are from a continuous population with distribution function $F(x)$ and probability density function $f(x)$, the joint probability density function for $X^{\mathbf{R}}_{1:m:n:k}, X^{\mathbf{R}}_{2:m:n:k}, \cdots, X^{\mathbf{R}}_{m:m:n:k}$ is given by

$$f_{1,2,\cdots,m}\left(x^{\mathbf{R}}_{1:m:n:k}, x^{\mathbf{R}}_{2:m:n:k}, \cdots, x^{\mathbf{R}}_{m:m:n:k}\right) = ck^m \prod_{j=1}^{m} f\left(x^{\mathbf{R}}_{j:m:n:k}\right)\left(1-F\left(x^{\mathbf{R}}_{j:m:n:k}\right)\right)^{k(\mathbf{R}_j+1)-1}, \quad (3)$$

$$0 < x^{\mathbf{R}}_{1:m:n:k} < x^{\mathbf{R}}_{2:m:n:k} < \cdots < x^{\mathbf{R}}_{m:m:n:k} < \infty, \quad (4)$$

where

$$c = n(n-R_1-1)(n-R_1-R_2-1)\cdots(n-R_1-R_2-\cdots-R_{m-1}-m+1). \quad (5)$$

There are four special cases:

The first one if $\mathbf{R} = (0,\cdots,0)$, Equation (3) reduces to the joint probability density function of first-failure-censored order statistics. The second case if $k=1$, Equation (3) becomes the joint probability density function of the progressively type II censored statistics. The third case if $k=1$ and $\mathbf{R}=(0,\cdots,0)$, then $n=m$ which corresponds to the complete sample. The last one if $k=1$ and $\mathbf{R}=(0,\cdots,n-m)$, then type II censored order statistics are obtained.

Also it can be seen that $X^{\mathbf{R}}_{1:m:n:k}, X^{\mathbf{R}}_{i:m:n:k}, \cdots, X^{\mathbf{R}}_{m:m:n:k}$ can be viewed as a progressively type II censored sample from a population with distribution function $1-(1-F(x))^k$. For this reason, results for the progressively type II censored order statistics can be extended to progressively first-failure-censored order statistics easily.

Obviously, although more items are used (only m of $n \times k$ items are failures) in the progressive first-failure censoring plan than in others, it has advantages in terms of reducing test cost and test time.

3. Point Estimation

In many cases, there will be an obvious or natural candidate for a point estimator of a particular parameter. For example, the sample mean is a natural candidate for a point estimator of the population mean. In this section, we estimate θ and p, by considering maximum likelihood, bootstrap and Bayes estimates. In Bayesian technique, we consider both symmetric (Squares Error, SE) loss function and asymmetric (Linear Exponential, LINEX and General Entropy, GE) loss functions.

3.1. Maximum Likelihood Estimation (MLE)

Let $x^{\mathbf{R}}_{i:m:n:k}, i=1,2,\cdots,m$, be the progressively first-failure censored order statistics from a Burr-X distribution, with censoring scheme $\mathbf{R}$ from (3), the likelihood function is given by

$$L(\theta;\underline{x}) = ck^m T(\underline{x})\theta^m \sum_{J_1=0}^{k(R_1+1)-1} \cdots \sum_{J_m=0}^{k(R_m+1)-1} G \exp\left(\theta \sum_{i=1}^{m}(j_i+1)\ln U_i\right), \quad (6)$$

where

$$T(\underline{x}) = \prod_{i=1}^{m} 2x_i \exp\left[-x_i^2\right](U_i)^{-1}, \quad U_i = \left(1-\exp\left\{-x_i^2\right\}\right), \quad (7)$$

$$G = (-1)^{j_1+\cdots+j_m}\binom{k(R_1+1)-1}{j_1}\cdots\binom{k(R_m+1)-1}{j_m}, \quad (8)$$

where c is defined in (5) and x_i is used instead of $x^{\mathbf{R}}_{i:m:n:k}$. Now, suppose that any group (r_i) being removed from the life test is independent of the others but with the same probability p. Then, the number of groups removed at each failure time follows a binomial distribution with parameters $n-m-\sum_{l=1}^{i-1} r_l$ and p where m is predetermined before the testing. Therefore

$$P(R_1=r_1)=\binom{n-m}{r_1}p^{r_1}(1-p)^{n-m-r_1}, \quad (9)$$

and for $i=2,3,\cdots,m-1$.

$$P(R_i=r_i|R_{i-1}=r_{i-1},\cdots,R_1=r_1) = \binom{n-m-\sum_{l=1}^{i-1} r_l}{r_i}p^{r_i}(1-p)^{n-m-\sum_{l=1}^{i-1} r_l}, \quad (10)$$

where $0\le r_i\le n-m-\sum_{l=1}^{i-1} r_l,\ i=2,3,\cdots,m-1$

$$P(R=r) = P(R_1=r_1)P(R_2=r_2|R_1=r_1)P(R_3=r_3|R_2=r_2,R_1=r_1) \cdots P(R_{m-1}=r_{m-1}|R_{m-2}=r_{m-2},\cdots,R_1=r_1), \quad (11)$$

hence

$$P(R=r)=\frac{(n-m)!}{\prod_{i=1}^{m-1} r_i!\left(n-m-\sum_{i=1}^{m-1} r_i\right)!} \times p^{\sum_{i=1}^{m-1} r_i}(1-p)^{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)r_i}. \quad (12)$$

Suppose further that R_i is independent of x_i. Then the likelihood function takes the following form

$$L(\theta,p;\underline{x},R)=L(\theta;\underline{x}|R=r)P(R=r), \quad (13)$$

$$L(\theta,p;\underline{x},r) = c^*T(\underline{x})\theta^m\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} G\exp\left(\theta\sum_{i=1}^{m}(j_i+1)\ln U_i\right) \times p^{\sum_{i=1}^{m-1} r_i}(1-p)^{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)r_i}. \quad (14)$$

Using (6), (12) and (13) we can write the likelihood function as

$$L(\theta,p;\underline{x},r)=c^*T(\underline{x})L_1(\theta;\underline{x})L_2(R=r), \quad (15)$$

where $c^*=\dfrac{ck^m(n-m)!}{\prod_{i=1}^{m-1} r_i!\left(n-m-\sum_{i=1}^{m-1} r_i\right)!}$ and

$$L_1(\theta;\underline{x})=\theta^m\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} G\exp\left(\theta\sum_{i=1}^{m}(j_i+1)\ln U_i\right). \quad (16)$$

and

$$L_2(R=r)=p^{\sum_{i=1}^{m-1} r_i}(1-p)^{\left((m-1)(n-m)-\sum_{i=1}^{m-1}(m-i)r_i\right)}. \quad (17)$$

It is obvious that L_1 in Equation (16) does not involve p. Thus the maximum likelihood estimate (MLE) of θ can be derived by maximizing Equation (16) directly. On the other hand, L_2 in Equation (17) does not depend on the parameter θ, then the MLE of p can be obtained directly by maximizing Equation (17). In particular, after taking the logarithms of $L_1(\theta)$ and $L_2(p)$, the MLE's of θ and p can be found by solving the following equations

$$\frac{\partial\log L_1(\theta)}{\partial\theta}=\frac{m}{\theta}+\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} G\sum_{i=1}^{m}(j_i+1)\ln U_i=0 \quad (18)$$

$$\frac{\partial\log L_2(p)}{\partial p} =\frac{\sum_{i=1}^{m-1} r_i}{p}-\frac{\left((m-1)(n-m)-\sum_{i=1}^{m-1}(m-i-1)r_i\right)}{1-p}=0. \quad (19)$$

Thus we find

$$\hat{\theta}_{ML}=(-m)\left(\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} G\sum_{i=1}^{m}(j_i+1)\ln U_i\right)^{-1}, \quad (20)$$

and

$$\hat{p}_{ML}=\frac{\sum_{i=1}^{m-1} r_i}{(m-1)(n-m)-\sum_{i=1}^{m-1}(m-i-1)r_i}. \quad (21)$$

3.2. Bootstrap Confidence Intervals

In this subsection, we use the parametric bootstrap percentile method suggested by Efron [13] to construct confidence intervals for the parameters. The following steps are followed to obtain a progressive first-failure censoring bootstrap sample from Burr-X distribution with parameter θ and binomial distribution with parameter p based on simulated progressively first-failure censored data with random removals set.

- From the original data $\underline{x}=x^{\mathbf{R}}_{1:m:n:k},x^{\mathbf{R}}_{2:m:n:k},\cdots,x^{\mathbf{R}}_{m:m:n:k}$ and $R=R_1,R_2,\cdots,R_m$ compute the ML estimates of the parameters $\hat{\theta}$ and $\hat{p}$ by Equations (20) and (21).
- Use $\hat{\theta}$ and $\hat{p}$ to generate a bootstrap sample $\underline{x}^*$ and R^* with the same values of $R_i,m;(i=1,2,\cdots,m)$ using algorithm presented in Balakrishnan and Sandhu [14] with distribution func-
- tion $\left[1-(1-F(x))^k\right]$ see Wu and Kuş [2].
- As in Step 1, based on $\underline{x}^*$ compute the bootstrap sample estimates of θ and p, say $\hat{\theta}^*$ and $\hat{p}^*$.
- Repeat steps 2-3 N times representing N bootstrap MLE's of (θ,p) based on N different bootstrap samples.
- Arrange all $\hat{\theta}^{*\prime}s$ and $\hat{p}^{*\prime}s$, in an ascending order to obtain the bootstrap sample $\left(\varphi_l^{[1]},\varphi_l^{[2]},\cdots,\varphi_l^{[N]}\right)$,

$l = 1,2$ (where $\varphi_1 \equiv \hat{\theta}^*, \varphi_2 \equiv \hat{p}^*$).

Let $G(z) = P(\varphi_l \le z)$ be the cumulative distribution function of φ_1. Define $\varphi_{lboot} = G^{-1}(z)$ for given z. The approximate bootstrap $100(1-\delta)\%$ confidence interval of φ_l is given by

$$\left[\varphi_{lboot}\left(\frac{\delta}{2}\right), \varphi_{lboot}\left(\frac{1-\delta}{2}\right)\right]. \tag{22}$$

3.3. Bayes Estimation

The Bayesian inference procedures have been developed under the usual SE loss function (quadratic loss), which is symmetrical, and associates equal importance to the losses due to overestimation and underestimation of equal magnitude. However, such a restriction may be impractical. For example, in the estimation of reliability and failure rate functions, an overestimation is usually much more serious than an underestimation; in this case the use of asymmetrical loss function might be inappropriate, as has been recognized by Basu and Ebrahimi [15], and Canfield [16].

A useful asymmetric loss known as the LINEX loss function, was introduced by Zimmer *et al.* [17], and was widely used in several papers by Balasooriya and Balakrishnan [18], Soliman [19] and Soliman [20]. This function rises approximately exponentially on one side of zero, and approximately linearly on the other side. Under the assumption that the minimal loss occurs at $\hat{u} = u$, the LINEX loss function for $u = u(\beta, \alpha)$ can be expressed as

$$L_1(\Delta) \propto \mathrm{e}^{a\Delta} - a\Delta - 1,\ a \ne 0, \tag{23}$$

where $\Delta = (\hat{u} - u)$, $\hat{u}$ is an estimate of u.

The sign, and magnitude of (a) represent the direction, and degree of symmetry. ($a > 0$ means overestimation is more serious than underestimation, and $a < 0$ means the opposite). For (a) closed to zero, the LINEX loss function is approximately the Squared Error (SE) loss, and therefore almost symmetric

The posterior-expectation of the LINEX loss function of (23) is

$$\begin{aligned} &E_u\left(L(\hat{u}-u)\right) \propto \exp(a\hat{u}) \\ &\times E_u\left[\exp(-au)\right] - a\cdot\left(\hat{u} - E_u[u]\right) - 1. \end{aligned} \tag{24}$$

where E_u is equivalent to the posterior-expectation with respect to the posterior $\mathrm{pdf}(u)$. The Bayes estimator $\hat{u}_{BL}$ of u under the LINEX loss function is the value $\hat{u}$, which minimizes (24)

$$\hat{u}_{BL} = -\frac{1}{a}\log\left(E_u\left[\exp(-au)\right]\right), \tag{25}$$

provided that $E_u\left[\exp(-au)\right]$ exists, and is finite.

Another useful asymmetric loss function is General Entropy (GE) loss

$$L_2(\hat{u}, u) \propto \left(\frac{\hat{u}}{u}\right)^b - b\log\left(\frac{\hat{u}}{u}\right) - 1. \tag{26}$$

whose minimum occurs at $\hat{u} = u$. This loss function is a generalization of the Entropy-loss used in several papers where $b = 1$ by Dey *et al.* [21], Dey and Liu [22]. When $b > 0$, a positive error $(\hat{u} > u)$ causes more serious consequences than a negative error. The Bayes estimate $\hat{u}_{BL}$ of u under GE loss (26) is

$$\hat{u}_{BG} = \left(E_u\left[u^{-b}\right]\right)^{(-1/b)} \tag{27}$$

provided that $E_u\left[u^{-b}\right]$ exists, and is finite.

Now, we assume that the parameters θ and p behave as independent random variables, and we use gamma prior distribution with known parameters α, β for θ. The prior pdf of θ takes the form

$$\pi_1(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{(\alpha-1)}\exp(-\beta\theta),\ \theta > 0, \alpha, \beta > 0, \tag{28}$$

while p has Beta prior distribution with known parameters γ, λ. That is

$$\pi_2(p) = \frac{1}{B(\gamma,\lambda)}p^{(\gamma-1)}(1-p)^{(\lambda-1)},\ 0 < p < 1,\ \gamma, \lambda > 0. \tag{29}$$

Therefore the posterior (pdf) of θ is

$$\pi_1^*(\theta|\underline{x}) = \frac{L_1(\theta;\underline{x})\pi_1(\theta)}{\int_0^\infty L_1(\theta;\underline{x})\pi_1(\theta)\mathrm{d}\theta}, \tag{30}$$

where $L_1(\theta;\underline{x})$ the likelihood function and $\pi_1(\theta)$ the prior density function. Applying (16) and (28), the marginal posterior (pdf) of θ given by

$$\pi_1^*(\theta|\underline{x}) = \frac{\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} G\theta^{(m+\alpha-1)}\exp(-\theta q_j)}{\Gamma(m+\alpha)\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1} Gq_j^{-(m+\alpha)}}, \tag{31}$$

where

$$q_j = \beta - \sum_{i=1}^m (j_i + 1)\ln U_i, \tag{32}$$

We notes that the posterior distribution of θ is Gamma with parameters $(m+\alpha)$ and q_j Similarly, the posterior (pdf) of p is

$$\pi_2^*(p|\underline{x}, r) = \frac{L_2(R=r)\pi_2(p)}{\int_0^1 L_2(R=r)\pi_2(p)\mathrm{d}p}, \tag{33}$$

where $L_2(R=r)$ the likelihood function and $\pi_2(p)$ the prior density function.

Applying (17) and (29), the marginal posterior pdf of p given by

$$\pi_2^*\left(p|\underline{x},r\right)=\frac{1}{B\left(\gamma^*,\lambda^*\right)}p^{\left(\gamma^*-1\right)}\left(1-p\right)^{\left(\lambda^*-1\right)}, \quad (34)$$

where

$$\gamma^*=\gamma+\sum_{l=1}^{m-1}r_i,$$
$$\lambda^*=\lambda+\left(m-1\right)\left(n-m\right)-\sum_{i=1}^{m}\left(m-i\right)r_i. \quad (35)$$

We notes that the posterior distribution of p is Beta with parameters γ^* and λ^*.

3.3.1. Symmetric Bayes Estimation

SE loss function: Under SE loss function, the estimator of a parameter (or given function of the parameters) is its posterior mean. Thus, Bayes estimators of the parameters are obtained by using the posterior densities (31) and (34). The Bayes estimators $\hat{\theta}_{BS}$ and $\hat{p}_{BS}$ of the parameters θ and p are

$$\hat{\theta}_{BS}=E_\theta\left(\theta|\underline{x}\right)=\int_0^\infty\theta\pi_1^*\left(\theta|\underline{x}\right)\mathrm{d}\theta \quad (36)$$

from (31) resulting in

$$\hat{\theta}_{BS}=\frac{\left(m+\alpha\right)\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}Gq_j^{-(m+\alpha+1)}}{\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}Gq_j^{-(m+\alpha)}}, \quad (37)$$

where G and q_j are defined in (8) and (32). Similarly,

$$\hat{p}_{BS}=E_p\left(p|\underline{x}\right)=\int_0^1 p\pi_2^*\left(p|\underline{x},r\right)\mathrm{d}p, \quad (38)$$

from (34) resulting in

$$\hat{p}_{BS}=\frac{\gamma^*}{\gamma^*+\lambda^*}, \quad (39)$$

where γ^* and λ^* are defined in (35).

3.3.2. Asymmetric Bayes Estimation

LINEX loss function: If in (25), $u=\theta$, then the Bayes estimator $\hat{\theta}_{BL}$, of the parameter θ relative to LINEX loss function is

$$\hat{\theta}_{BL}=-\frac{1}{a}\log\left[\int_0^\infty\exp\left(-a\theta\right)\pi_1^*\left(\theta|\underline{x}\right)\mathrm{d}\theta\right], \quad (40)$$

and from (31), we get

$$\hat{\theta}_{BL}=-\frac{1}{a}\log\left[\frac{\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}G\left(a+q_j\right)^{-(m+\alpha)}}{\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}Gq_j^{-(m+\alpha)}}\right]. \quad (41)$$

Similarly, if in (25), $u=p$, then the Bayes estimator $\hat{p}_{BL}$, of the parameter p relative to LINEX loss function is

$$\hat{p}_{BL}=-\frac{1}{a}\log\left[\int_0^1\exp\left(-ap\right)\pi_2^*\left(p|x\right)\mathrm{d}p\right], \quad (42)$$

and from (34), we obtain

$$\hat{p}_{BL}=-\frac{1}{a}\log\left[\frac{1}{B\left(\gamma^*,\lambda^*\right)}\int_0^1\exp\left(-ap\right)p^{\left(\gamma^*-1\right)}\left(1-p\right)^{\left(\lambda^*-1\right)}\mathrm{d}p\right]. \quad (43)$$

One can use a numerical integration technique to get the integration in (43).

General Entropy loss function: Let $u=\theta$, in (27), then the Bayes estimate $\hat{\theta}_{BG}$, of parameter θ relative to the General Entropy loss function is

$$\hat{\theta}_{BG}=\left[\int_0^\infty\theta^{-b}\pi_1^*\left(\theta|\underline{x}\right)\mathrm{d}\theta\right]^{(-1/b)}, \quad (44)$$

and from (31), we obtain

$$\hat{\theta}_{BG}=\left[\frac{\Gamma\left(m+\alpha-b\right)\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}Gq_j^{-(m+\alpha-b)}}{\Gamma\left(m+\alpha\right)\sum_{j_1=0}^{k(R_1+1)-1}\cdots\sum_{j_m=0}^{k(R_m+1)-1}Gq_j^{-(m+\alpha)}}\right]^{(-1/b)}. \quad (45)$$

Put $u=p$ in (27), then the Bayes estimator $\hat{p}_{BG}$ of the parameter p relative to General Entropy loss function is

$$\hat{p}_{BG}=\left[\int_0^1 p^{-b}\pi_2^*\left(p|\underline{x}\right)\mathrm{d}p\right]^{(-1/b)}, \quad (46)$$

from (34), resulting in

$$\hat{p}_{BG}=\left[\frac{\Gamma\left(\gamma^*+\lambda^*\right)\Gamma\left(\gamma^*-b\right)}{\Gamma\left(\gamma^*\right)\Gamma\left(\gamma^*+\lambda^*-b\right)}\right]^{(-1/b)}. \quad (47)$$

4. Interval Estimation

4.1. Approximate Interval Estimation

The asymptotic variances and covariances of the MLE for parameters θ, and p are given by elements of the inverse of the Fisher information matrix

$$\boldsymbol{I}_{ij}=E\left[-\frac{\partial^2 L}{\partial\theta\partial p}\right];\ i,j=1,2. \quad (48)$$

Unfortunately, the exact mathematical expressions for the above expectations are very difficult to obtain. Therefore, we give the approximate (observed) asymptotic varaince-covariance matrix for the MLE, which is obtained by dropping the expectation operator E

$$\begin{bmatrix} -\frac{\partial^2 L}{\partial \theta^2} & -\frac{\partial^2 L}{\partial \theta \partial p} \\ -\frac{\partial^2 L}{\partial p \partial \theta} & -\frac{\partial^2 L}{\partial p^2} \end{bmatrix}^{-1}_{(\hat{\theta},\hat{p})} = \begin{bmatrix} \mathrm{var}(\hat{\theta}) & 0 \\ 0 & \mathrm{var}(\hat{p}) \end{bmatrix}, \quad (49)$$

with

$$\frac{\partial^2 L}{\partial \theta^2} = -\frac{m}{\theta^2}, \quad (50)$$

$$\frac{\partial^2 L}{\partial p^2} = \frac{\sum_{i=1}^{m-1} r_i}{p^2} + \frac{(m-1)(n-m) - \sum_{i=1}^{m-1}(m-i) r_i}{(1-p)^2}. \quad (51)$$

The asymptotic normality of the MLE can be used to compute the approximate confidence intervals for parameters θ and p. Therefore, $(1-\delta)100\%$ confidence intervals for parameters θ and p become

$$\hat{\theta} \pm Z_{\delta/2}\sqrt{\mathrm{var}(\hat{\theta})} \quad \text{and} \quad \hat{p} \pm Z_{\delta/2}\sqrt{\mathrm{var}(\hat{p})}, \quad (52)$$

where $Z_{\delta/2}$ is the percentile of the standard normal distribution with right-tail probability $\delta/2$.

4.2. Highest Posterior Density Interval (HPDI)

In general, the Bayesian interval estimation is much more direct than frquentest classical method. Now, having obtained the posterior distribution $p(\omega|Data)$ we ask, "How likely is it that the parameter ω lies within the specified interval $[\omega_L, \omega_U]$?" Bayesian call this interval based on the posterior distribution a "credible interval". The interval $[\omega_L, \omega_U]$ is said to be a $(1-\delta)100\%$ credible interval for θ if

$$\int_{\omega_L}^{\omega_U} p(\omega|Data)\mathrm{d}\omega = 1-\delta. \quad (53)$$

For the shortest credible interval, we have to minimize the interval $[\omega_L, \omega_U]$ subject to the condition (53) which requires

$$P(\omega_L|Data) = P(\omega_U|Data). \quad (54)$$

As interval $[\omega_L, \omega_U]$ which simultaneously satisfies (53) and (54) is called the "shortest" $(1-\delta)100\%$ credible interval. A highest posterior density interval (HPDI) is such that the posterior density for every point inside the interval is greater than that for every point outside of it. For a unimodal, but not necessarily symmetrical posterior density, the shortest credible and the HPD intervals are identical. We now proceed to obtain the $(1-\delta)$ 100% HPD intervals for the parameters θ and p. Consider the posterior distribution of θ in (31). The $(1-\delta)100\%$ HPDI $[\theta_L, \theta_U]$ for the parameter θ is given by the simultaneous solution of the equations

$$\int_{\theta_L}^{\theta_U} \pi_1^*(\theta|\underline{x}, r) = (1-\delta) \quad \text{and} \quad \pi_1^*(\theta_L|\underline{x}, r) = \pi_1^*(\theta_U|\underline{x}, r). \quad (55)$$

Similarly, using the posterior pdf of p in (34), the $(1-\delta)100\%$ HPDI $[p_L, p_U]$ for the parameter p is given by the simultaneous solution of the equations

$$\int_{p_L}^{p_U} \pi_2^*(p|\underline{x}, r) = (1-\delta) \quad \text{and} \quad \pi_2^*(p_L|\underline{x}, r) = \pi_2^*(p_U|\underline{x}, r). \quad (56)$$

To obtain the HPDI from (55) and (56), one may employ any mathematical package such as Mathematica, to get the intervals.

5. Numerical Example

Example 1: (*simulated data*) To illustrate the use of the estimation methods proposed in this paper. A set of data consisting of 75 observations were generated from a Burr-X distribution with parameter $\theta = 0.79$, and randomly grouped into 15 sets. The generated data are listed below:

0.3194	0.4661	0.8348	0.1150	0.1230	0.2136	0.1373	0.2053
0.7253	0.7738	0.9407	0.1516	0.5111	0.4148	0.1599	0.3227
0.9233	0.8316	0.9615	0.2006	0.9758	0.6618	0.4116	0.9088
0.9787	0.8461	0.9795	0.7353	1.2692	0.8632	0.4503	1.1337
0.9956	0.9732	1.2067	1.0935	1.8144	0.9698	1.0344	2.0543
0.1775	0.3165	0.2732	0.2832	0.2752	0.2814	0.2761	
0.3363	0.7871	0.4714	0.6613	0.5764	0.7273	0.5616	
0.5353	0.8052	0.5134	0.6790	0.7441	0.7602	0.6529	
0.8312	0.8695	0.9356	0.9049	0.9939	1.2178	0.9328	
1.9019	1.0115	1.1291	1.5136	1.2023	1.7956	1.2521	

Now, we consider the following cases:

Case I: Progressive first-failure censored data with binomial removals.

Algorithm 1.

1) Specify the value of n.

2) Specify the value of m

3) Generate the value of the parameters θ and p using the prior densities (28) and (29), for some given values of the prior parameters α, β, γ and λ.

4) Generate a random number r_1 from $bio(n-m, p)$.

5) Generate a random numbers r_i from

$$bio\left(n-m-\sum_{l=1}^{i-1} r_i, p\right), \quad \text{for each } i, \quad i = 2, 3, \cdots, m-1.$$

6) Set r_m according to the following relation,

$$r_m = \begin{cases} n-m-\sum_{l=1}^{i-1} r_l & \text{if } n-m-\sum_{l=1}^{i-1} r_l > 0 \\ 0 & \text{o.w} \end{cases}$$

Based on the above data, a progressive first-failure censored data with binomial removals were generated using the algorithm described in Balakrishnan and Sandhu [14] with distribution function $\left[1-\left(1-F(x)\right)^k\right]$ see Wu and Kuş [2].

The generated progressive first-failure censored data with binomial removals are: (0.115, 0.123, 0.1373, 0.1757, 0.2053, 0.2732, 0.2752, 0.2761, 0.2832, 0.4661), and R = (0, 3, 1, 0, 1, 0, 0, 0, 0, 0) Using the results presented in previous sections, the different point estimates of θ and p are computed. We denote to the MLEs, estimates using the bootstrap, Bayes estimate relative to SE loss, Bayes estimate relative to LINEX loss, and Bayes estimate relative to GE loss, respectively by $(.)_{ML}, (.)_{Boot.}, (.)_{BS}, (.)_{BL}$ and $(.)_{BG}$. The results are displayed in **Table 1**.

Case II: First-failure censoring data with n ($n=15, m=15$ and $k=5$).

The set of the first-failure censored data are: (0.115, 0.123, 0.1373, 0.1757, 0.2053, 0.2136, 0.2732, 0.2752, 0.2761, 0.2814, 0.2832, 0.3165, 0.3194, 0.4661, 0.8348). Different point estimates of θ are computed and the results are listed in **Table 1**.

Case III: Progressive type II censoring data with binomial removals.

A progressive type II censoring data with binomial removals have been generated from complete sample using the algorithm described in Balakrishnan and Sandhu [14], with ($n=75, m=50, k=1, p=0.4$ and $R=\left(9,8,2,3,1,0,0,2,0^{42}\right)$). *i.e.* 50 failure times are observed and 25 failure times are censored using censored scheme R. The generated data are:

(0.115, 0.123, 0.1516, 0.1599, 0.2006, 0.2053, 0.2136, 0.2752, 0.2761, 0.2814, 0.2832, 0.3165, 0.3194, 0.3227, 0.3363, 0.4116, 0.4148, 0.5111, 0.5134, 0.5616, 0.5764, 0.6529, 0.679, 0.7273, 0.7353, 0.7441, 0.7602, 0.7871, 0.8052, 0.8312, 0.8461, 0.8632, 0.8695, 0.9049, 0.9088, 0.9328, 0.9407, 0.9698, 0.9732, 0.9787, 0.9939, 0.9956, 1.0344, 1.0935, 1.1291, 1.2067, 1.2178, 1.5136, 1.7956, 1.8144). The results of different Bayes estimates of θ and p are also, listed in **Table 1**.

Case IV: The complete sample data with ($n=m=75$ and $k=1$)

The results of point estimates of the parameter θ and p are shown in **Table 1**.

Based on different type of censoring described above, the 95% credible intervals of θ and p are obtained using approximate confidence interval (ACI), confidence interval based on bootstrap re-sampling method (Boot CI), and the highest posterior density interval (HPDI). All the results are listed in **Table 2**.

Table 1. Different point estimates of θ and p for all cases with $(\theta, p)=(0.79, 0.4)$.

					$(.)_{BL}$		$(.)_{BG}$	
Cases	(.)	$(.)_{ML}$	$(.)_{Boot.}$	$(.)_{BS}$	$a=-2$	$a=2$	$b=-2$	$b=2$
I	θ	0.7724	0.7863	0.7697	0.7823	0.7578	0.7776	0.7463
	p	0.3571	0.3974	0.3529	0.3659	0.3405	0.3705	0.2887
II	θ	0.7818	0.7936	0.7775	0.7885	0.7670	0.7844	0.7570
III	θ	0.7656	0.7723	0.7608	0.7701	0.7518	0.7667	0.7428
	p	0.3906	0.3985	0.3881	0.3916	0.3846	0.3925	0.3740
IV	θ	0.7802	0.7902	0.7745	0.7825	0.7667	0.7796	0.7592

Table 2. 95% confidence intervals for θ and p under progressive first-failure censored samples when $(\theta, p)=(0.79, 0.4)$.

Cases	(.)	ACI	Boot CI	HPDI
		[L,U]	[L,U]	[L,U]
I	θ	[0.5524, 0.9925]	[0.6965, 0.8056]	[0.5737, 1.0069]
	p	[0.1061, 0.6081]	[0.1905, 0.5000]	[0.1520, 0.5866]
II	θ	[0.3862, 1.1775]	[0.5285, 0.6365]	[0.5911, 0.9964]
III	θ	[0.5753, 0.9559]	[0.7015, 0.8076]	[0.5864, 0.9606]
	p	[0.2711, 0.5102]	[0.3906, 0.4902]	[0.2758, 0.5066]
IV	θ	[0.6036, 0.9568]	[0.6928, 0.8798]	[0.6102, 0.9581]

6. Simulation Study

In order to compare the different estimators of the parameters, we simulated 1000 progressively first-failure-censored samples from a Burr type X distribution with the values of parameters $(\theta, p)=(0.79, 0.4)$, and different combinations of n, m, k, and censoring random schemes **R**. The samples were simulated by using the algorithm described in Balakrishnan and Sandhu [14]. A simulation was conducted in order to study the properties and compare the performance of the Bayes estimator with maximum likelihood estimator.

The mean square error (MSE) of the Bayes estimations and maximum likelihood estimations are computed over different combination of the censored random scheme **R** as shown in **Tables 3** and **4**. To asses the effect of the shape parameters *a* and *b* from **Tables 3** and **4**, one can see that the asymmetric Bayes estimates (*BL*, *BG*) of the (MSE) of the parameters θ and p are overestimates for ($a<0$, $b<0$), and when ($a>0$, $b>0$) the (MSE) of the parameters are underestimates. Also, the MSE of Bayes estimates are smaller than MSE of the MLE, when $(b=-1)$ the MSE of Bayes estimates relative to *GE* loss are the same as the MSE relative to SE loss Bayes estimates. As anticipated, all MSE of Bayes estimates

Table 3. Mean square errors of the parameter θ.

K	n	m	ML	BS	BL a −1	BL a 1	BG b −1	BG b 1
1	30	15	0.1729	0.1561	0.1625	0.1512	0.1561	0.1526
		20	0.1590	0.1440	0.1495	0.1398	0.1440	0.1412
		25	0.1574	0.1433	0.1484	0.1392	0.1433	0.1401
	50	35	0.1332	0.1248	0.1283	0.1218	0.1248	0.1213
		40	0.1270	0.1198	0.1226	0.1176	0.1198	0.1178
		45	0.1180	0.1115	0.1140	0.1095	0.1115	0.1096
3	30	15	0.1088	0.1044	0.1067	0.1026	0.1044	0.1025
		20	0.1032	0.0993	0.1012	0.0978	0.0993	0.0976
		25	0.1001	0.0966	0.0982	0.0951	0.0966	0.0949
	50	35	0.0829	0.0808	0.0819	0.0798	0.0808	0.0794
		40	0.0769	0.0754	0.0760	0.0749	0.0754	0.0751
		45	0.0743	0.0728	0.0735	0.0722	0.0728	0.0721
5	30	15	0.0937	0.0913	0.0926	0.0903	0.0913	0.0904
		20	0.0888	0.0866	0.0879	0.0856	0.0866	0.0854
		25	0.0832	0.0812	0.0823	0.0803	0.0812	0.0800
	50	35	0.0660	0.0651	0.0656	0.0648	0.0651	0.0649
		40	0.0641	0.0633	0.0637	0.0630	0.0633	0.0630
		45	0.0614	0.0606	0.0610	0.0604	0.0606	0.0604
7	30	15	0.082	0.0804	0.0813	0.0797	0.0804	0.0798
		20	0.0773	0.0760	0.0767	0.0755	0.0760	0.0576
		25	0.0722	0.0711	0.0716	0.0706	0.0711	0.0706
	50	35	0.0592	0.0585	0.0590	0.0582	0.0585	0.0580
		40	0.0549	0.0543	0.0546	0.0540	0.0543	0.0540
		45	0.0523	0.0519	0.0521	0.0518	0.0519	0.0519

Table 4. Mean square errors of the parameter p.

K	n	m	ML	BS	BL a −1	BL a 1	BG b −1	BG b 1
1	30	15	0.0865	0.0767	0.0779	0.0757	0.0767	0.0753
		20	0.1056	0.0886	0.0901	0.0873	0.0886	0.0882
		25	0.1518	0.1066	0.1098	0.1038	0.1066	0.1065
	50	35	0.0871	0.0772	0.0783	0.0763	0.0772	0.0761
		40	0.1065	0.0887	0.0904	0.0872	0.0887	0.0871
		45	0.1618	0.1112	0.1148	0.1080	0.1112	0.1089
3	30	15	0.0856	0.0765	0.0774	0.0757	0.0765	0.0762
		20	0.1052	0.0883	0.0899	0.0868	0.0883	0.0873
		25	0.1498	0.1047	0.1081	0.1016	0.1047	0.1033
	50	35	0.0835	0.0744	0.0754	0.0734	0.0744	0.0734
		40	0.1037	0.0874	0.0889	0.0859	0.0874	0.0864
		45	0.1464	0.1039	0.1068	0.1014	0.1039	0.1053
5	30	15	0.0879	0.0779	0.0790	0.0769	0.0779	0.0765
		20	0.1065	0.0894	0.0910	0.0880	0.0894	0.0883
		25	0.1512	0.1066	0.1098	0.1037	0.1066	0.1062
	50	35	0.0857	0.0766	0.0776	0.0757	0.0766	0.0758
		40	0.1095	0.0915	0.0932	0.0900	0.0915	0.0901
		45	0.1471	0.1031	0.1063	0.1004	0.1031	0.1036
7	30	15	0.0879	0.0780	0.0791	0.0771	0.0780	0.0770
		20	0.1070	0.0896	0.0912	0.0881	0.0896	0.0883
		25	0.1564	0.1097	0.1129	0.1068	0.1097	0.1092
	50	35	0.0861	0.0766	0.0776	0.0756	0.0766	0.0753
		40	0.1065	0.0890	0.0907	0.0874	0.0890	0.0872
		45	0.1546	0.1082	0.1115	0.1052	0.1082	0.1074

relative to both LINEX loss, and GE loss (for a close to 0, and $b=-1$) are the same as the SE loss Bayes estimates. This one of the useful properties of working with the LINEX loss function we found that for different choices of k, n, m and censoring random scheme $\mathbf{R}$ the MSE of the Bayes estimates based on symmetric and asymmetric loss functions perform better than MSE of the MLEs. when the effective sample proportion m/n increases, the MSE of each the Bayes estimation and maximum likelihood estimations reduce. Also the censoring scheme $R=(n-m,\cdots,0)$ is most efficient For all choices, it seems to usually provide the smallest MSE for each estimates of θ and p.

7. Conclusion

The purpose of this paper is to develop a Bayesian analysis for Burr-X distribution under the progressively first-fialure censored sampling plan with binomial random removals. We studied point and interval estimations of parameter of the Burr type X distribution. We derived the MLEs, Bayes estimators (BS, BL, BG). A simulation study was conducted to examine the performance of the MLE and the Bayes estimators.

8. Acknowledgements

The authors would like to express their thanks to the editor, assistant editor and referees for their useful comments and suggestions on the original version of this manuscript.

REFERENCES

[1] L. G. Johnson, "Theory and Technique of Variation Re-

search," Elsevier, Amsterdam, 1964.

[2] S.-J. Wu and C. Kus, "On Estimation Based on Progressive First-Failure-Censored Sampling," *Computational Statistics and Data Analysis*, Vol. 53, No. 10, 2009, pp. 3659-3670. doi:10.1016/j.csda.2009.03.010

[3] P. L. Gupta, S. Gupta and Ya. Lvin, "Analysis of Failure Time Data by Burr Distribution," *Communication Statistics Theory & Methods*, Vol. 25, No. 9, 1996, pp. 2013-2024. doi:10.1080/03610929608831817

[4] A. Childs and N. Balakrishnan, "Conditional Inference Procedures for the Laplace Distribution When the Observed Samples Are Progressively Censored," *Metrika*, Vol. 52, No. 3, 2000, pp. 253-265. doi:10.1007/s001840000092

[5] S. K. Tse, C. Y. Yang and H.-K. Yuen, "Statistical Analysis of Weibull Distribution Lifetime Data under Type II Progressive Censoring with Binomial Removals," *Journal of Applied Statistics*, Vol. 27, No. 8, 2000, pp. 1033-1043. doi:10.1080/02664760050173355

[6] M. A. M. Ali Mousa and Z. F. Jaheen, "Statistical Inference for the Burr Model Based on Progressively Censored Data," *An International Computers & Mathematics with Applications*, Vol. 43, No. 10-11, 2002, pp. 1441-1449. doi:10.1016/S0898-1221(02)00110-4

[7] K. Ng, P. S. Chan and N. Balakrishnan, "Estimation of Parameters from Progressively Censored Data Using an Algorithm," *Computational Statistics and Data Analysis*, Vol. 39, No. 4, 2002, pp. 371-386. doi:10.1016/S0167-9473(01)00091-3

[8] S.-J. Wu and C.-T. Chang, "Parameter Estimation Based on Exponential Progressive Type II Censored Data with Binomial Removals," *Information and Mangement Statistics*, Vol. 13, No. 3, 2002, pp. 37-46.

[9] N. Balakrishnan, N. Knnan, C. T. Lin and H. Ng, "Point and Interval Estimation for Gaussian Distribution Based on Progressively Type-II Censored Samples," *IEEE Transactions on Reliability*, Vol. 52, No. 3, 2003, pp. 90-95. doi:10.1109/TR.2002.805786

[10] S.-J. Wu, "Estimation for the Two-Parameter Pareto Distribution under Progressive Censoring with Uniform Removals," *Journal of Statistical Computation and Simulation*, Vol. 73, No. 2, 2003, pp. 125-134. doi:10.1080/00949650215732

[11] A. A. Soliman, "Estimation of Parameters of Life from Progressively Censored Data Using Burr-XII Model," *IEEE Transactions on Reliability*, Vol. 54, No. 1, 2005, pp. 34-42. doi:10.1109/TR.2004.842528

[12] A. M. Sarhan and A. Abuammoh, "Statistical Inference Using Progressively Type-II Censored Data with Random Scheme," *International Mathematical Forum*, Vol. 35, No. 3, 2008, pp. 1713-1725.

[13] B. Efron, "The Bootstrap and Other Resampling Plans," *CBMS-NSF Regional Conference Seriesin Applied Mathematics*, Vol. 38, SIAM, Philadelphia, 1982.

[14] N. Balakrishnan and R. Asandhu, "A Simple Simulation Algorithm for Generating Progressively Type-II Censored Samples," *The American Statistican*, Vol. 49, No. 2, 1995, pp. 229-230.

[15] A. P. Basu and N. Ebrahimi, "Bayesian Approach to Life Testing and Reliability Estimation Using Asymmetric Loss Function," *Journal of Statistical Planning and Inference*, Vol. 29, No. 1-2, 1991, pp. 21-31. doi:10.1016/0378-3758(92)90118-C

[16] R. V. Caneld, "A Bayesian Approach to Reliability Estimation Using a Loss Function," *IEEE Transactions on Reliability*, Vol. 19, No. 1, 1970, pp. 13-16.

[17] W. J. Zimmer, J. Bert Keats and F. K. Wang, "The Burr XII Distribution in Reliability Analysis," *Journal of Quality & Technology*, Vol. 30, No. 4, 1998, pp. 386-394.

[18] U. Balasooriya and N. Balakrishnan, "Reliability Sampling Plans for Log-Normal Distribution Based on Progressively Censored Samples," *IEEE Transactions on Reliability*, Vol. 49, No. 2, 2000, pp. 199-203. doi:10.1109/24.877338

[19] A. A. Soliman, "Comparison of LINEX and Quadratic Bayes Estimators for the Rayleigh Distribution," *Communications in Statistics Theory and Methods*, Vol. 29, No. 1, 2000, pp. 95-107. doi:10.1080/03610920008832471

[20] A. A. Soliman, "Estimation of Parameters of Life from Progressively Censored Data Using Burr-XII Model," *IEEE Transactions on Reliability*, Vol. 54, No. 1, 2005, pp. 34-42.

[21] D. K. Dey, M. Ghosh and C. Srinivasan, "Simultaneous Estimation of Parameters under Entropy Loss," *Journal of Statistical Planning and Inference*, Vol. 15, No. 2, 1987, pp. 347-363.

[22] D. K. Dey and P. L. Liu, "On Comparison of Estimation in Generalized Life Model," *Microelectron & Reliability*, Vol. 32, No. 1-2, 1992, pp. 207-221. doi:10.1016/0026-2714(92)90099-7

Intelligent Information Management, **2013, 5, 126-135**
doi:10.4236/iim.2013.54013 Published Online July 2013 (http://www.scirp.org/journal/iim)

Scientific Research

General Frame Work of New TQM Based on the ISO/IEC25000 Series of Standard

Kazuhiro Esaki
Faculty of Science and Engineering, Housei University, Kajino-cho, Koganei-shi, Tokyo, Japan
Email: Kees959@hotmail.com

Received April 8, 2013; revised May 15, 2013; accepted June 17, 2013

ABSTRACT

Generally, for quality improvement of an organization management, TQM (Total Quality Management) is used worldwide and recognized. However, contents of activity are various, and it is difficult for organization to define the whole scope of TQM. It is very important that the activity theme about the TQM is based on the consideration of the needs and priority from the objective requirement, which should correspond to the basic consideration of quality management. If we have selected the wrong theme, and missed understanding about importance and priority of target for TQM, it may cause a significant risk in the organization management. For example, the company which got a high evaluation about the quality of management has poor management and disgraceful affair after several years, and it is a case reaching the bankruptcy. We worked on a study for improvement of technique for quality requirement and evaluation of system for many years, and have participated in the development of international standard for supporting the quality requirements and evaluation of system in ISO/IECJTC1/SC7. The organization having a specific purpose is a kind of system. It is thought that evaluation technique of the system quality is adaptive for a quality evaluation of the organization management. Therefore, this study suggests the new framework of TQM from the view point of the general framework of management defined in ISO/IEC25040 and TQM matrix. Furthermore, this paper presents the result of comparison between proposed action theme derived from a framework of new TQM and the example of existing TQM, and reports an inspection result of the effectiveness.

Keywords: TQM; Structure of Quality; TQM Matrix; Quality Assurance; Quality Improvement; Investment Management; Project Management; Static Risk Management; Dynamic Risk Management

1. Introduction

Generally, for quality improvement of an organization management, TQM (Total Quality Management) is used worldwide and recognized. The TQM has spread in USA in the 1990s, and to provide a high quality and suitable price of product and service at appropriate good timing for a customer satisfaction. In order to provide the best practice, TQM is the strategic system-like methodology to contribute to achieve the purpose of organization effectively. Not only one department but also the companies (a manager, a director, a worker, and others) and activity of QC are extended to "a design section, a production section, a purchasing section, a sales section, a marketing section and maintenance or service".

It is the system for company-wide quality control activities. The biggest aim is covering all processes from a plan, the design stage of the product to production, sale, maintenance, service and manages the quality of the product generally, and it is to aim at the effective and high quality management. As for the TQC, the principal objective is put for the quality improvement of the work process including the application to the indirect section whereas QC is an action coherent to the production spot.

TQM is the thing which improved the TQC more, and the characteristic for the activity is strategic and the top management should commit the total quality management and should lead top-down. The TQM clarifies requirements for the management quality of the organization by the examination standard for ex, American Malcolm Baldridge Prize [1] and Japanese management quality prize [2] etc. now. However, contents of activity are various, and it is difficult for organization to define the whole scope of TQM. It is very important that the activity theme about the TQM is based on the consideration of the needs and priority from the objective requirement, which should correspond to the basic consideration of quality management. If we have defined the wrong theme, and miss understanding about importance and priority of target for TQM, as a result, it may cause a significant risk in the organization management.

For example, the company which got a high evaluation about the quality of management causes poor management and disgraceful affair after several years, and there is the case reaching the bankruptcy. Such a situation has the risk to give many stakeholders concerned including the stockholder who misjudged a vast loss.

We worked on a study for improvement of quality requirement and evaluation of system for many years, and have participated in the development of international standard for supporting the quality requirements and evaluation of system in ISO/IECJTC1/SC7.

In recent years, we have been working on developing the ISO/IEC25000 (SQuaRE) series [3-8] of standards for quality requirements and evaluation for system and software product for a long time in ISO/IEC JTC1 (Joint Technical Committee 1 of the International Organization for Standardization and the International Electro technical Commission) SC7WG 6 (software and systems engineering under ISO technical committee, working group six).

As part of this project, we have worked on the developments of ISO/IEC25030, 25040, 25041 [6-8], which are the standards to provide supporting technology for the method of quantitative quality requirement definitions [6] and evaluation [7,8] based on the ISO/IEC9126-1 [5,9] quality model is widely recognized and used in world for the purpose of specify the quality requirement and evaluation of system/software product. On the other hand, definition of system is "a combination of interacting elements organized to achieve one or more stated purposes" defined in ISO/IEC15288: 2008 Systems and software engineering—System life cycle processes [10]. The organization having also a specific purpose is a kind of system, too. It is thought that evaluation technique of the system quality is adaptive for a quality evaluation of the organization management.

Therefore, this study suggest the new framework of TQM from the view point of the general framework of management defined in both proposed TQM matrix and elevation framework of ISO/IEC25040. Furthermore, this paper present the result of comparison between proposed activity theme derived from a framework of new TQM and the example of existing TQM, and report an inspection result of the effectiveness of proposed total framework of new TQM.

2. Concept

2.1. Concept of Quality Improvement

Figure 1 shows the concept of quality control that show the relationships between process and product, and concept of the PDC cycle. From **Figure 1**, every activity includes a product and a process. In addition, an input and output of process is included in the activity of organizations. Also, the quality control activity makes the repetition of product and process, which include a "product-process-product-process and product" during a PDC cycle as shown in **Figure 1**. Output product of a process before becomes an input product for the later process. All kind of management processes is affected by an input quality of own process, which is a result of previous management process. Therefore, the quality of processes depends on a quality of previous process, and not to be able to exist alone each.

—Definition of Quality

From ISO, Terminology "Quality" is defined as "totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs". Objects of a quality control usually are quality of product and quality of process. Furthermore, quality has both primary quality and second quality.

—PDC Cycle

It is necessary to turn the PDC cycle called the inspection to improve the quality of product and process to show in **Figure 1**. It should visualize a quality state of a product and process for turn a PDC cycle, and it is necessary to confirm the existence of problems. Generally, PDC is called PDCA as Plan-Do-Check-Action.

And, improvement activity should be performed immediately after checking and resolve the problems.

But "action" of improvement is same as "Do".

Therefore some kind of plan should be studied in front of the action because it is thought necessary. There is the risk that improvement activity leads to a change for the worse when we do not perform some kind of plan before improvement actions. Therefore, in this study, "PDCA" introduced "PDC".

—Primary Quality

It is to meet the requirements for the function that

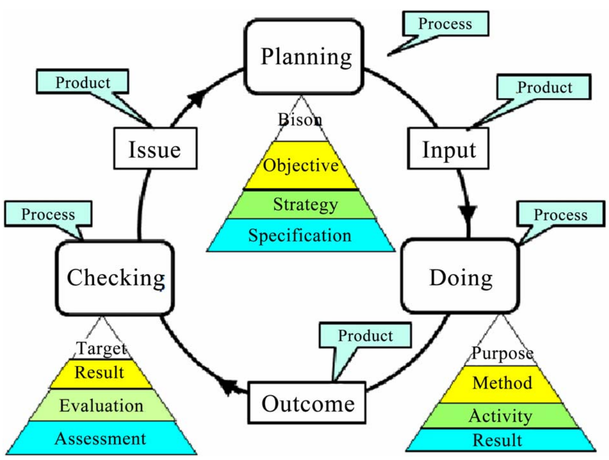

Figure 1. Concept of quality control.

should be realized specified in a contract and a product specification, or it is to meet the business processes prescribed in quality or an international standard and a rule in the company. The un-satisfaction for primary quality requirement is a problem, and it is thought that the problem can be finally zero if we push forward correction essentially.

—Secondary Quality

It is an attractive quality to meet the tacit expectation of a customer needs. And the quality that we solve a problem having, and can be accomplished now. The issue is that a product and a process do not meet secondary quality. The problem never disappears as far as environment surrounding an organization changes. Therefore, it is necessary the solution to problem soaks a priority based on the need in order to be able to go ahead.

—Product Quality

The product is a product or service, result brought about as a result of input resources and activity of the organization activity, and there are the primary quality, secondary quality and quality in use.

When we define the quality of system product, the model of six quality characteristics of "functionality, reliability, efficiency, usability, portability, maintainability defined in ISO/IEC9126 and used a system product.

We show the influence that a product gives to a user with four quality characteristics such as "effectiveness, productivity, safety, satisfaction" as included in quality in use. "Quality at the time of the use" is quality of the use effect when it used a product in specific use purpose, environment, and it is the quality that a user finally feels by a product and a conformity degree of the use environment.

—Process Quality

The process is the process that converts input resources into the output, and it is a procedure or a method of the activity. We can regard the quality of process as efficiency to convert an input product in the organization activity into an output product. In ISO9001, are the requirements for the process followed with the quality of the process? Does a PDC cycle rotate?

We evaluate the attribute of the process and judge it.

Even if we measure the quality of process only, we cannot judge the quality of process. The quality of the process after the improvement is shown in a lower equation.

—Efficiency of process = quality of outcome/input resource.

And evaluation of improvement degree of the process, it is necessary to evaluate that how improved for the quality of input resources essentially.

2.2. Concept of the Structure of Quality

Figure 2 shows the concept of "structure of quality problem and issue" by suggested in this study.

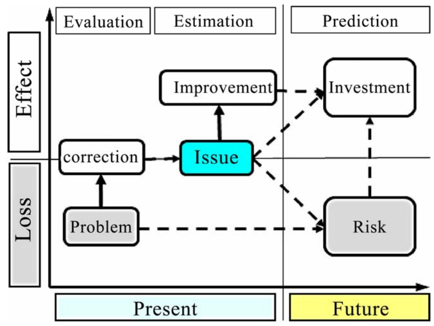

Figure 2. Concept of structure of quality.

The cross axle shows the present and the future in time axes, and the vertical axis shows a volume of positive effect or negative effect (a loss) by the influence of occurred problem.

It is necessary to correct problem in a normal condition immediately because the problem (the primary quality incompatible) that occurred in a product or process is in condition to have deviated from a plan and specifications during past or at present. And a negative effect (a loss) occurs when it take the influence of the quality in time, a vertical axis on the cross axle to show it in **Figure 2**.

As shown in **Figure 2**, the problem may cause a negative effect at present, and a more serious negative effect (a loss) will be at risk of the minus number to be connected in it. On the other hand, an issue (deficiency of the second quality) that occurred in a product or process is deviation from an attractive requirement actualized in the past or at present, and there is not the negative effect (a loss), but positive because it is not the deviation from a plan and specifications. Therefore, it is necessary to perform the solution based on the consideration of priority or importance and urgent degree of problem.

As shown in **Figure 2**, issue may cause big problem in future even if the issue does not have a negative effect (a loss) at present. On the other hand, the issue is actualized as a positive effect in the future when we make the improvement plan for expected positive effect and perform it. In this way, risk and investment management can be a necessary theme for TQM that is not a scope of quality control base on the structure of quality problem, and it is thought that should be intended theme for TQM.

—Problem

The problem is abnormal condition to have deviated from the normal condition should be. Following are example of problems.

—Trouble, an error, a defect, an obstacle (fault, obstacle), not conform.

—Problem of Product

Result of activity cannot realize the required plan or specifications with the deficiency of an actualized product.

—**Problem of Process**

It is that PDC does not rotate with the deficiency of an actualized process. The over budget that are not pushed forward in a planned schedule, delivery delay, incompatible. For example, it is not to meet the requirements of ISO9000 and the inner company rules.

—**Issue**

It is the actualized refinement that may bring improvement of the quality of a product and the process.

The outbreak probability of a certain phenomenon means the combination of results of the phenomenon.

—**Risk**

Probability of phenomenon and size of the damage when problem will occur.

2.3. Concept of General Management Framework for Organization

Figure 3 is the simplest concept of the organization activity that paid its attention to the specific process and product of the concept of the quality control that showed in **Figure 1** and it is defined from the general framework of quality requirement and evaluation for system included in ISO/IEC25040. Furthermore, this framework is based on a concept of IDEF0 [11].

Figure 3 shows the product of four elements of new TQM and the management of the process.

In this study, the framework of the total quality management for improvement of the organizations quality is that the "input for, outcomes of, constrain for, own resources" and "the conversion process from input to outcomes" should included as shown in **Figure 3**.

The organization activity is regarded as a process to convert some kind of products (input resources) into some kind of products (Outcomes). In this study, "input, outcomes and resources" is called product, and "conversion procedure" is called process. It is important to control the organization activity effectively for quality control forward. And, organization should produce high quality product (value and a product, service) from the view point of possible limited input and resources of organization own. In this study, I devised the theme of the organization management necessary to lead an organization to the success from such a viewpoint.

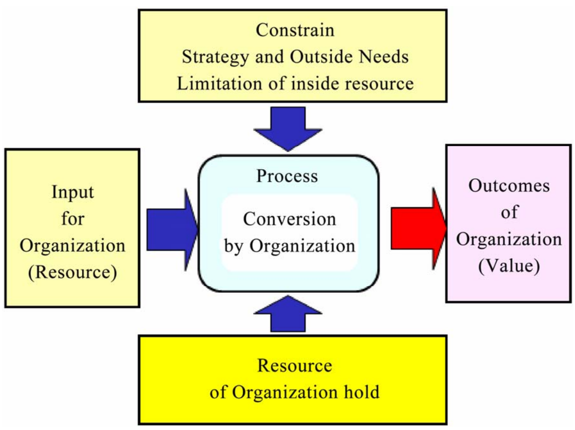

Figure 3. Frame work of quality evaluation. -ISO/IEC25040 (based on the IDEF0) [11].

3. Concept of the Total Quality Management Matrix

Figure 4 shows the more detail and devised domain of new TQM framework, which has introduced based on the structure of quality problem as explained in **Figure 2**.

The time of the cross axle of **Figure 2** can sort with the outbreak probability of the phenomenon more finely as shown in **Figure 4**. On the other hand, an effect and loss can define more detail in the vertical axis on a scale from the view point of influence of problem.

Based on the consideration of **Figure 4**, matrix of the whole object domain of the total quality management can be defined.

The domain of the total quality management has covered both quality assurance and quality improvement conventionally that has been intended. But, additional management such as investment (project management) and the risk management should cover from the view point of consideration of **Figure 4**. **Figure 4** shows the four necessary domain of management for new TQM.

3.1. Area of Quality Assurance

—**Domain of the Quality Assurance.**

Activities of assurance of the primary quality of product and service that are specified and promised according to the contract of customers at past and present, and provide to customers. The activity of guarantee of quality corrects a problem shown in **Figure 4**, and should secure quality of normal state. For example, problem outbreak is such as violating of the contract, laws and ordinances or

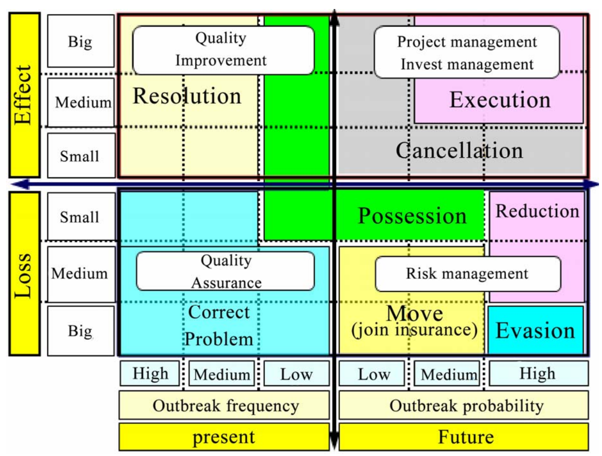

Figure 4. Concept of TQM matrix.

a rule in the company. The activity to guarantee the quality of the products such as intermediate or final product and process such as the deficiency of the procedure of the specified plan, and aim of contract of customer should be achieved. Quality of output product and process should be achieved with the promised cost and the appointed date of delivery.

If result of activity could not come true according to the contract and the project plan, the problem should be corrected immediately. It is usually necessary to correct because the guarantee of quality responsibility (defect responsibility) based on the contract may cause a problem that occurred. And, it is necessary to correct problem promptly, because the problem has produces the negative effect (a loss). Correspondence of second quality is an improvement of the problems for the realization of attractive quality to satisfy the tacit expectation of the customer and is not included in a quality assurance.

If we cannot guarantee the primary quality of the product, the problem may occur after delivery to customer. When the worst, a recall may occur and asked the compensation for damages from a customer and continuation of the business is spoiled by loss of the trust of the organization. Therefore, a quality assurance is the required activities that perform surely in the organization. In addition, it is necessary to spread it horizontally so that not only correct a problem of merely activity but also avoid the same problem occurs in other projects and other organizations. Furthermore, provisional measures and permanent problem should be divided and should correct them. It is necessary to divide it into pieces for measures by consideration of the seriousness of the problem and an urgent degree, the skill of the personnel required, cost, the limitation of the period in provisional measures and permanency, and to carry it out.

3.2. Area of Quality Improvement

—Domain of the Quality Improvement

It is a domain of the quality improvement such as the QC (Quality Control) circles which have been performed conventionally. That activity solves the issue that an organization holds, and to raise it in attractive quality of product of the organization and the second quality of the process, and to meet an expected requirement. The issue does not produce the negative effect at that point, but may cause positive effect, if we can improve it. We cope the improvement issue actualized in the past or at present that an attractive quality of product or the more effective procedures. The issues may be changed by the application environment of the result.

And it is necessary to consider the priority while ascertaining advancement of the aim, reduction in cost, strength and emergency degree of the demand of customer including shortening on the appointed delivery date because defect responsibility does not necessarily occur, and to cope. We can hit a priority in consideration of the importance, emergency degree and can go ahead through the improvement only when we evaluate the effect that we can expect when we solve the issue by the improvement. The remaining problems are limited as far as there is a contract with customers, but issue may not always disappear. The improvement of the secondary quality is an everlasting theme of the quality control.

3.3. Area of Static Risk Management

—Domain of Static Risk Management

It is a domain called risk management conventionally.

It may cause future problem by the influenced from a product (the intermediate, final) or the process (the deficiency of the requirement procedure) of the present organization and is generally called as risk management.

Risk is defined as the scale of the damage and probability that a problem will produce in future when occurred.

It may cause big problem for organization in future, if we leave a present hold problem when do not take any improvement actions and may lay a negative effect (a loss). When we do not hit any measures and improve the present conditions, it is a problem that will occur in future by the turn of inside situation or change of outside environment of organization, and this case has called "static risk" in this study. We investigate a risk to be inherent in a product and a process of the present based on correspondence of policy, risk analysis to a static risk and, by risk assessment, calculate the outbreak probability and scale of the damage at the time of the outbreak quantitatively and estimate it. In addition, as for the risk measures, it is necessary performing based on the consideration of the priority from the effectiveness and limited input resources, and to perform "reception, imputation, reduction and evasion" four risk measures correspond to the result of risk analysis as the TQM matrix as shown in **Figure 4**.

3.4. Area of Investment Management

—Domain of the Investment Management

The domain of the quality of the investment activity, it is a domain called the project management or investment management conventionally. The management of the investment risk is generally handled as risk management of the projects out of an object of the quality management in PMBOK [12]. The project can be called the investment activity to get a positive effect to expect in the future. Therefore, the project management can be called the management of the investment risk, because risk may be caused by the activity of project.

Investment management should be included in the

TQM, because that should achieve a current quality problem or issues for taking the expected positive effect (value) in the future. In this study, the degree of the “positive or negative effect” that is caused by the influence from project when we hit the measures expecting improvement in future, investment management risk can be called dynamic risk. From **Figure 4**, the dynamic risk is defined on the “probability and scale of an effect” when it occurred by the project which produce an expectation or negative effect by taking new action for improvement. Inherent risk of project for a purpose and the accomplishment of a project described in the plan should be estimated at the planning before project start.

From **Figure 4**, it is necessary for the risk management to perform judgment such as enforcement or avoidance from the view point of success probability and the cost-effectiveness in response to the result of risk and the portfolio analysis as a general rule.

4. Total Quality Management Framework

Figure 5 is the overall framework of the new TQM which is introduced from a concept of both **Figures 3** and **4** suggested. From **Figure 5**, “input resources, outcomes and each domain of management process” has systematized. Furthermore, **Figure 5** shows the framework of whole TQM, which include such as the input for, outcomes, constrains and a supporting resources for organization activity and conversion process from input into the outcomes. The risk management of project and investment management have been handled as the specific management theme and out of scope of TQM.

But, in this study, management of risk of project and investment can be the important theme of TQM, and should be included in the domain of TQM from the consideration of TQM matrix as shown in **Figure 4**.

From **Figure 5**, input for the organizations may be the demand from various stakeholders which occurred in the

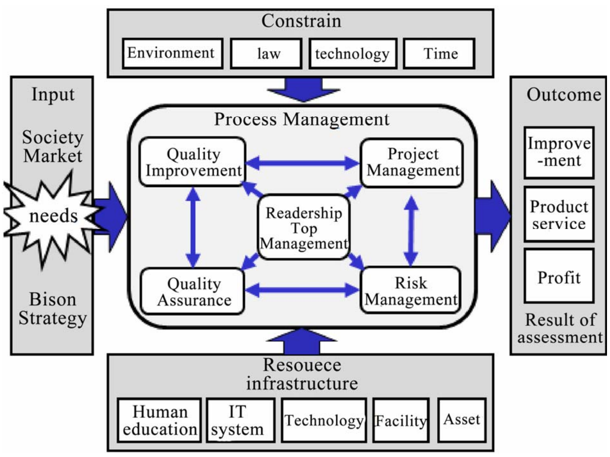

Figure 5. Total frame work of TQM.

external environment of the organization. Furthermore, the outcome is the result of organization activity.

The organization activity is carried out with input resources in the line of the factory and the specific environment of work, and as the result, product and service has be provided. In addition, as the secondary effect, it produces various improvement effects or brings the negative effect except organization purposes such as the discharge of the CO_2. The improvement includes the wide meaning for various for products such as input and outcomes. And discharge of CO_2 can be recognized as the negative improvement effect.

First, constrain is the demand for the organizations that include aim, budgets and time limit of which agreed with customers. Second, constrain include the limitation such as law, rule, standard, corporate strategy of the organization, financial resource, human resource, facilities environment and technologies to support the organization activities. The resource of organization is the supporting infrastructures that include the various management technology, financial resources, information system and human resources etc.

In this study, I have proposed the new framework of TQM, that is necessary to lead an organization for success of improvement from such a viewpoint and considerations.

4.1. Management of Input

The input is the new resources need to adopt from the outside necessary for organization activity as follows.

—Demand from customer, human resources, facilities, materials, engineering, technique necessary for the demand from the customers, etc.

—Demand from Customer

It is grounds to determine the necessity of the organization activity, and the grasp of the demand from customer is important to identify a purpose of the organization strategy management and the plan management. The demand is the time limit of activity, customer’s expectation of quality and quantity of the cost, etc. Demand of customer is the both agreed “demand stated clearly” and “the potential demand” that is the “implicit demand”. We should grasp it through a requirements definition process with customer. And the necessary demand from customer should be defined in to incorporate the aim of organization planning appropriately.

—Human Resources

The human resources are the main subject of the organization activity. Quality improvement and the efficiency of the organization activity depend on the quality of human resources greatly. The human resources is the only existence having the thing true to make a value, and it is an organization comprised by the personnel required and workers. In addition, the organization is formed by the human resources based on some purpose of existence ideas. Human resources should have the ability and specialty that adapted to the every process of organization management from the demanded. For the purpose of achievement of organization successfully, procurement of the human resources should be performed based on constrain from organization strategy.

4.2. Management of Outcomes

Quality management is the activity for the quality of result of "product or service, and improvement of process" by performing TQM for the purpose of improvement as shown in **Figure 5**.

—Economical Result

The output of the process is the result of "profit, sales, cost reduction, cash flow and asset" by total quality management activity as shown in **Figure 5**. The information such as profitability or productivity, cause or measures that occurred by a process of activity of organization are included.

—Result of the Organization Activity

The output of the process is the result of "quality of product or service, and improvement of process" by total quality management activity as shown in **Figure 5**.

The information such as problems or issues, cause or measures that occurred by a process of activity of organization are included. The result of process such as "product quality, appointed date of delivery and the cost improvement defined in the plan as aim" may be achieved finally by using these quality related information. It is the aim of activity that should be accomplished by tissue, and the aim should be achieved.

The product should be popular in the world, and value is broken into by you, and advantage is provided.

The result of quality in use may be different in response to a demand of customer's needs. The result of application effect may the result of "direct or indirect improvement" provided by the organization. The characteristic of direct result is the characteristics of new product developed and information system introduced, etc.

The characteristic of effect is the improved work process of organization, the expansion of the sales provided by the injection of new product to the market, result of improvement of productivity and business efficiency based on the installed new information system, improvement of the profit, etc.

—Effect of Improvement (Kaizen)

An object of the improvement has the outcome (a product or service) and the process of organization.

The effect of improvement is the degree of improvement of human skill, the quality of input and outcomes, result of the process improvement of organization.

4.3. Management of Constrain

—Constrain of the Organization Activity

It is led by "laws, Corporate strategy, inner companies laws, regulation, demand" from the outside of organization. Also led by the limitation of resources such as "financial resources, human resources, facilities, material, and technique" of organization. It is the limitation included in an aim set to an organization plan. And the example of elements are the "purpose, delivery date, planned budget of project" and strategy of organization, etc.

—Strategy of Organization

They have the inseparable relationship between corporate strategy and the process activities of organization.

It is necessary to draft the organization strategy of the organization as a premise of the organization activity.

The organization activity is placed in the low rank of the corporate strategy and influenced by the limitation of corporate strategy.

—Financial Resources of Organization

The organization activity has the limitation by financial resources of organization. For example, the human resources of company have an influence on the success or failure of the organization activity very much. And, lack of human resources may cause the failure.

—Technique of Organization

The organization activity has the limitation of technology of organization. For example, technique, method, tools, management standard, management style of supporting, management rule and a standard procedure, etc.

The limitation of the management support technology has an influence on the success or failure of the organization activity very much. And, lack of technology may be big limitation of the organization activity.

—Laws and Ordinances, Regulation

Laws and ordinances are applied to control organization activity. And, such elements are internal law and rule, a product standard, an international standard, the office rule that organization activity should obey. ISO9000 and the Labor Standards Law is recognizes applied law and rule, and it is necessary to obey it surely.

4.4. Management of Resources

It is the support base which an organization helping practice of the organization activity holds. Examples of elements are information system, financial resources, human resources, facilities, material and core technology, etc.

—Financial Resources Which an Organization Holds

The organization activity that a company organization carries out is affected by the financial resources which the organization holds. For example, the human resources that a company holds have an influence on the suc-

Table 1. Comparison of total quality management.

Proposed Framework of new TQM			Example of conventional TQM	
Management Theme		Content	Baldrige Criteria for Performance Excellence	Japan Quality Award
Product	Input control	Management resources to acquire for a purpose from the outside of the organization *Information *A finance *Human resources *Facility, equipment *New technology& Tools	Customer Focus (Category 3) 3.1 Voice of the Customer	Understanding of Customer & Market —Make confidence with customer —Define the customer needs —Select and analysis of information system —Select and analysis of management information
	Outcomes control	Result of the organization activity *product *Service *Profit Improvement of product & Process *Information system *Finance *Human resources **Facility *Possession technology& Tools *Organization	1.2 Governance and Societal Responsibilities Results (Category 7) 7.1 Product and Process Results 7.2 Customer-Focused Results 7.3 Workforce-Focused Results Purpose 7.5 Financial and Market Results	Result of activity —Result of corporate responsibility —Result of leadership —Result of improvement of organization —Result of improvement of human skill —Result of process —Result of finance —Result of customer satisfaction
	Constrain control	A strategy and an organization rule for limitation from the outside for purposes and control product Needs from customer *Quality Requirement *Schedule & Budget *Possession technology& Tools *Organization	1.2 Governance and Societal Responsibilities Strategic Planning (Category 2) 2.1 Strategy Development 2.2 Strategy Implementation	Strategic Planning and adaptation —Strategy Development —Strategic control-Select and analysis of management information
	Resource control	The reinforcement of the management base which an organization holds *Inner company Rule *A support system *A financing base *Human resources *Education *A possession technology& Tools *Knowledge data base	4.2 Knowledge Management, Information, and Information Technology, Workforce Focus (Category 5) 5.1 Workforce Environment 5.2 Workforce Engagement	Promotion process of Value —Supporting process —Select and analysis of information system —Improvement information system —Improvement of organization & human resource —Improvement of human skill —Education of human resource —Improvement of satisfaction of personnel —Improvement of working environment
Process	Top Management Readership	The leadership and the responsibility of the person in charge of the organization *Setting of the composition target *Decision making *Strategic control *Accountability *corporate responsibility	Leadership (Category 1) 1.1 Senior Leadership 1.2 Governance and Societal Responsibilities Strategic Planning (Category 2) 2.1 Strategy Development 2.2 Strategy Implementation	Leadership of top management Social responsibility of top management —Correspondent of social needs —Contribution of society-Strategic control —Select and analysis of management information
	Quality Assurance	Work on a guarantee of quality *product *process	2.2 Strategy Implementation Operations Focus (Category 6) 6.1 Work Processes	—Result of customer satisfaction

Continued

Quality Improvement	Quality improvement activities *A product A person, a thing, money, information *A process A procedure, a method, rules	Measurement, Analysis, and Knowledge Management (Category 4) 4.1 Measurement, Analysis, and Improvement of Organizational Performance 4.2 Knowledge Management, Information, and Information Technology Operations Focus (Category 6) 6.1 Work Processes 6.2 Operational Effectiveness	Process of Value promotion —Business process improvement —make confidence with customer
Risk management	Static risk management to reduce a negative effect	----	----
Investment Project management	Dynamic risk management in hope of a positive effect *Investment management *Project management *Finance investment	----	----

cess or failure of the organization activity very much.

And human resources and the facilities become the big support factor of the organization activity.

—The Technique That an Organization Holds

A theory and the technique are utilized for the purpose of quality improvement and efficiency of the organization activity. A technique for the organization management is necessary to let an organization succeed.

The possession of technology becomes the big support factor of the organization activity. It is necessary to identify a necessary thing from the plan stage of the organization, and to introduce it to vary according to the process of the organization management.

—Support Information System

The organization activity is supported by introduced information system.

5. Verification

Table 1 shows the result of comparison between evaluation criteria of new TQM that has proposed in this study and evaluation criteria of other conventional TQM.

From **Table 1**, quality control of forward investment activity and the risk management for the future is thin in the conventional TQM is recognized.

Furthermore, it is thought that "the management of input, outcomes, constrain and resource" is weak based on the result of **Table 1** from the view point of new TQM framework proposed in this study.

Input resources limited as for the good TQM is given, and, as a result, under the limitation condition, how much value can provide? TQM is the relative quality of efficiency to say how much improvement you can bring about. And, for the purpose of true improvement of organization management quality, not only improve process management but also improvement of "input, outcomes, constrain and resource of the organization" of organization should be performed.

6. Conclusions

In this study, I suggested the theme that it was necessary to really perform it in TQM from the viewpoint of the framework of the new TQM for an activity theme of the TQM which had been pushed forward conventionally as shown in **Figure 5**. From the result of the comparison of **Table 1**, the activity of forward investment and the risk management for the future seems to be short in the conventional TQM. Therefore, we should cope with the spatial and time turn of events of the outside of organization when you would like to provide suitable for the sustained development of the organization and will offer the continuous stability-like value to a customer.

Based on a framework of new TQM, it is thought that the reinforcement of the action for forward investment activity (like project management or innovation) and the risk management that stood in the long term prospects is necessary as well as a short term.

In addition, in the TQM, we should manage the improvement of quality of "input, outcomes, constrain, resources", and it is thought with the theme that is extremely important to the reinforcement of activity improving. And it is thought that it is important to evaluate how much you laid the result from limited input resources. In the framework of new TQM shown in **Figure 5**, I put the leadership of top management in the center of **Figure 5** and locate quality assurance, quality improvement, risk management, investment management in the circumference. It is thought that mind about values of the top leader is very important located in the center of the TQM.

The leadership of top management is extremely important for the improvement of management quality.

If the quality of top management is bad even if the management system of organization is good, we cannot continue to supply the good outcome such as product or service. As a result, it is thought that we cannot secure the consequent profit.

In the future study, the author plans to study the evaluation indicators of management quality of organization by each theme proposed in this study and would like to verify of them.

7. Acknowledgements

The authors are grateful to members of production system research office at Graduate School of Factory of Science and Engineering HOSEI University who their contributions and support to make the discussion.

REFERENCES

[1] American Malcolm Baldridge Prize, Criteria for Performance Excellence, 2014. http://www.nist.gov/baldrige/publications/business_nonprofit_criteria.cfm/

[2] Japanese Management Quality Prize, Criteria for Performance Excellence. http://www.jpc-net.jp/eng/award/

[3] ISO/IEC25000, "Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)—Guide to SQuaRE," International Organization for Standardization, 2005.

[4] ISO/IEC25001, "Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)—Planning and Management," International Organization for Standardization, 2007.

[5] ISO/IEC25010, "Software Engineering-System and Software Quality Requirements and Evaluation (SQuaRE)—System and Software Quality Model," International Organization for Standardization, 2011.

[6] ISO/IEC25030, "Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)—Quality Requirement," International Organization for Standardization, 2007.

[7] ISO/IEC25040, "Software Engineering-System and Software Quality Requirements and Evaluation (SQuaRE)—Evaluation Process," 2011.

[8] ISO/IEC25041: "Software Engineering-System and Software Quality Requirements and Evaluation (SQuaRE)—Evaluation Guide for Developers, Acquirers and Independent Evaluators," 2012.

[9] ISO/IEC9126-1, "Software Engineering-Product Quality —Part1: Quality Model," 2001.

[10] ISO/IEC15288, "Systems and Software Engineering System life Cycle Processes—System Life Cycle Processes," International Organization for Standardization, 2008.

[11] Federal Information Processing Standards Publication 183, Integration definition for Function Modeling (IDEF0), 1993.

[12] Project Management Body of Knowledge, 4th Edition, Project Management Institute, 2008.

Intelligent Information Management, **2013, 5, 136-139**
doi:10.4236/iim.2013.54014 Published Online July 2013 (http://www.scirp.org/journal/iim)

A Hybrid Algorithm for Stemming of Nepali Text

Chiranjibi Sitaula
Central Department of Computer Science and Information Technology, Tribhuvan University, Kathmandu, Nepal
Email: candsbro@gmail.com

Received April 25, 2013; revised May 26, 2013; accepted June 15, 2013

ABSTRACT

In this paper, a new context free stemmer is proposed which consists of the combination of traditional rule based system with string similarity approach. This algorithm can be called as hybrid algorithm. It is language dependent algorithm. Context free stemmer means that stemmer which stems the word that is not based on the context *i.e.*, for every context such rule is applied. After stripping the words using traditional context free rule based approach, it may over stem or under stem the inflected words which are overcome by applying string similarity function of dynamic programming. For measuring the string similarity function, edit distance is used. The stripped inflected word is compared with the words stored in a text database available. That word having minimum distance is taken as the substitution of the stripped inflected word which leads to the stem of it. The concept of traditional rule based system and corpus based approach is heavily used in this approach. This algorithm is tested for Nepali Language which is based on Devanagari Script. The approach has given better result in comparison to traditional rule based system particularly for Nepali Language only. The total accuracy of this hybrid algorithm is 70.10% whereas the total accuracy of traditional rule based system is 68.43%.

Keywords: String Similarity; Information Retrieval; Text Mining; Natural Language Processing; Dynamic Programming

1. Introduction

Stemming means finding the root or stem from the given inflected word. It is used in Natural Language processing, Information Retrieval, Text Mining etc. Mostly, stemming is used to improve the performance for NLP (Natural Language Processing). For example, if the word such as "उपरथिहरु" is used for NLP. Searching with this long string may degrade the performance, but if the stemming is done with this word *i.e.* रथि. Obviously, the performance is increased because we don't need to search other unnecessary words.

Apart from the Natural Language Processing Task, the stemming plays a very important role in text mining task of computer science. In the process of stemming normally the input tokens are given the core engine which strips the inflected word leading to proper root which is used for better searching in search engine (**Figure 1**).

For stemming purpose, different algorithms are available in text mining purpose. They include rule based, machine learning based and statistical based algorithms.

The input and output of the data are given in **Table 1** for the demonstration. The required output is given into second column whereas the word to be tested is given in the first column.

2. Literature Review

Several works have been performed in the field of stemming including German, Spanish, Indian and etc. Talking about Indian stemming which is more similar to Nepali

Table 1. Mapping of inflected word and stem/root.

Inflected word	Root
उपरथिहरु	रथि
समाचार	आचार
संवेदना	वेदना
औपन्यासिक	उपन्यास
अत्यावश्यक	आवश्यक

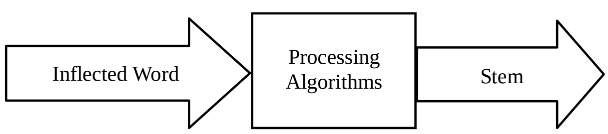

Figure 1. Processing for Stemmer.

stemming, have performed stemming work in their own different local languages like Tamil, Punjabi, Bengali, Gujarati, Hindi, Marathi etc. Similarly, Arabians have also performed such operation in their local language. Although many algorithms exists, they are mainly focused on their native language only.

[1] performed stemming approach in Arabic text. They used five methods. Of which four of these were positional letter ranking approach and fifth was traditional rule based system and found that rule based system performed well when combining with correction algorithm. Similarly, [2] used machine learning approach for stemming approach and it performed high accuracy. For performing such stemming approach they used different classifier like Naïve Bayesian, Bayesian Network, OneR, ZeroR and J148 algorithms. [3] used light stemmer with heuristic and co-occurrence of information retrieval approach. For handling co-occurrence they used clustering approach. [4] used different existing stemming algorithms and comparisons is made. They include Lovins, Porter Stemmer etc. [5] discussed about different stemming algorithms for English Language and a hybrid algorithm is made for Gujarati incorporating different algorithms. Some performed stemming of the Arabic text using Hidden Markov model [6] which gave more accuracy. Similarly [7] stemmed the Hindi Texts using hybrid approach. In this approach, they used the combination of brute force and suffix stripping approach which tries to remove the problem of over-stemming and under-stemming. The very preliminary phase of Hindi was Light Stemmer which was done by [8]. It is just a set few rules with the help of which stemming are done. Similarly, [9] uses stemming algorithms for Punjabi words which are based on the algorithms defined by [10,11]. It exploits brute force approach with few stripping strategy. It just matches the patterns and displays the root from the database. It the searching word is found, its respective stem is retrieved otherwise it just stems the affix and gives the output. [12] used stemming approach for text classification in Arabic language. For preprocessing the texts, they used stemming and performed classification. Similarly, [12-17] has done research on their different languages like Bengali, Punjabi etc. [18] performed research of different stemming algorithms. [19] explains stemming algorithms for Arabic language using parallel corpus and [20] has explained different stemming algorithms. [21-23] also performed different stemming approach for their own language like Nepali and Turkish.

3. Proposed Model

For explaining the proposed model, following things are taken into consideration: prefix, suffix and root.

3.1. Suffix

Suffix means those words that come after the root or stem. Around 150 suffixes are taken into consideration.

3.2. Prefix

These are the words that are added to the front part of the Nepali words. Around 35 words are taken.

3.3. Roots

Around 700 complex stem words are taken into consideration for research activity.

3.4. Hybrid Methodology

This algorithm is based on classical rule based stemming algorithms like [8]. It exploits the features of string similarity function of dynamic programming [23]. The complete flow of algorithm is given in **Figure 2** below.

This algorithms in context free algorithm. *i.e.*, it doesn't care about the context of the word. It strips the words from the inflected word. For example, in this algorithm, after entering the inflected word the stripping operation is performed. Incremental stripping approach is

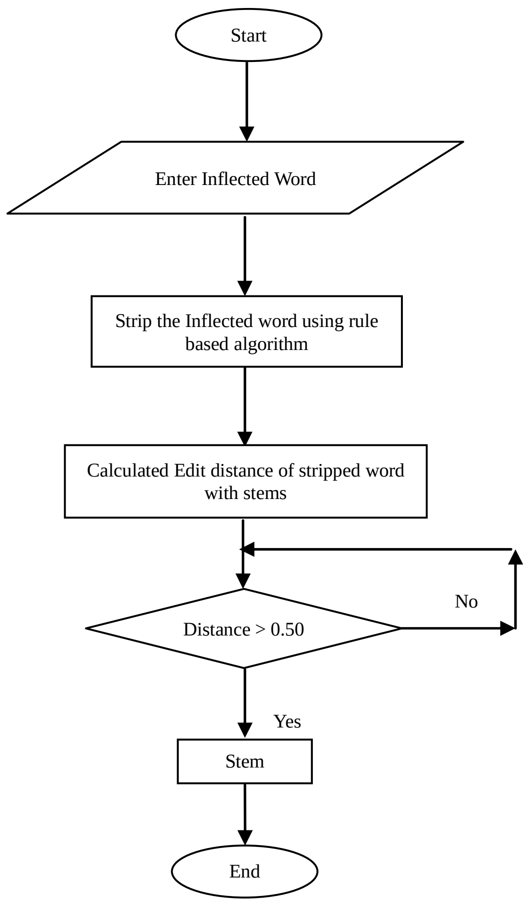

Figure 2. Flow of operation.

employed for suffix portion. But for prefix portion, it is not employed. For example, उपरथिहरुले = उप + रथि + हरु + ले. In this example, “हरु” and “ले” are two suffixes stripped through incremental approach and “उप” is prefix. Similarly, for prefix portion longest length stripping approach is followed. In this method, the longest size of the prefix is stripped although there may be presence of another affix. For example, अमान्छे = अ + मान्छे and अधिकरण = अधि + करण. In this example, “अ” and “अधि” both are prefix but priority is given to the prefix having longest size like “अधि” not “अ”.

As 0.50 was found to be the best threshold in [24], it is taken as the best threshold. After stripping the word, the words are compared with the roots stored in database using string similarity function which has used dynamic programming approach for comparing.

The output obtained from this stop is stripped word but in some time, the word may be over stripped or under stripped. In order to compensate the over stripped or under stripped words, the concept of string similarity approach is exploited.

4. Evaluation and Output

For the evaluation purpose, around 1200 complex words are taken as test keywords. It was implemented under Visual Studio 2008. Programming language was C#. For measuring the performance, precision and recall are used.

The comparison of the hybrid algorithms with traditional rule based algorithm is made. The output is listed in **Table 2**.

Similarly, the output of the stemming using traditional rule based system is listed in **Table 3**.

5. Conclusion and Limitation

After performing the research on stemming of Nepali Keywords, following conclusions are made:

- The recall of rule based system was 68.43 and the recall of Enhanced system was 72.1.
- The over stripping and under stripping are recovered by Enhanced System.
- Its context free nature is not handled.
- Few rules are applied.
- Incremental stripping of prefix is not allowed.
- Longest length stripping is not applied in suffix portion although it is applied in prefix portion.
- It can be compared with many other algorithms.
- The less number of words stored in corpus leads to wrong output so more number of words are necessary in corpus.
- Different thresholds for measuring the distance can be used.
- Different similarity measures can be used and compared.

Table 2. Output of rule based system.

Group	No. of keywords	Recall	Average recall
1	400	75.25	
2	400	74.45	72.1
3	497	67.60	

Table 3. Output of modified algorithms.

Group	No. of keywords	Recall	Average recall
1	400	74.25	
2	400	70.45	68.43
3	497	60.60	

6. Acknowledgements

Thanks to Dr. Bipul Shyam Purkayastha from Assam University, India for providing me an implausible support. Similarly, special thanks go to Mr. Bikash Balam from Central Department of Computer Science and Information Technology, Tribhuvan University along with my students and colleagues for supporting me.

REFERENCES

[1] Y. AI-Nashashibi, D. D. Neagu and Y. Ali, “Stemming Techniques for Arabic Words: A Comparative Study,” 2*nd International Conference on Computer Technology and Development* (*ICCTD*), 2010, pp. 270-276.

[2] H. Mohammad, B. Zuhair, C. Keely and M. David, “An Arabic Stemming Approach Using Machine Learning with Arabic Dialogue System,” *ICGST AIML*-11 *Conference*, Dubai, April 2011, pp. 9-16.

[3] L. S. Leah, B. Lisa and C. E. Margaret, “Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurance Analysis,” *SIGIR*, ACM, 11-15 August 2002.

[4] L. S. Leah, B. Lisa and C. E. Margaret, “Conservatice Stemming for Search and Indexing,” *ACM*, August 2005, pp. 15-19.

[5] S. Jikitsha and P. C. Bankim, “Stemming Techniques and Naïve Approach for Gujarati Stemmer,” *International Conference in Recent Trends in Information Technology and Computer Science*, IJCA, 2012, pp. 9-11.

[6] A. F. Alajmi, E. M. Saad and M. H. Awadalla, “Hidden Makov Model Based Arabic Morphological Analyzer,” *International Journal of Computer Engineering Research*, *IJCER*, Vol. 2, No. 2, 2011, pp. 28-33.

[7] M. Upendra and P. Chandra, “MAULIK: An Effective Stemmer for Hindi Lanuage,” *International Journal of Computer Science and Engineering*, *IJCSE*, Vol. 4, No. 5, 2012, pp. 711-717.

[8] R. Ananthakrishnana and R. D. Durgesh, “A Light Stemmer for Hindi.”

[9] K. Dinesh and R. Kumar, “Design and Development of

Stemmer for Pujabi," *International Journal of Computer Applications*, *IJCA*, Vol. 11, No. 12, 2010, pp. 18-23. doi:10.5120/1634-2196

[10] S. Llia, "Overview of Stemming Algorithms," Depaul University.

[11] F. B. William and F. J. Christopher, "Strength and Similarity of Affix Removal Stemming Algorithms," James Madison University and Virginia Tech.

[12] O. H. M. Ali and L. Ma Shi, "Stemming Algorithm to Classify Arabaic Documents," *Symposium on Progress in Information & Communication Technology*, 2009, pp. 111-115.

[13] A. James and K. Giridhar, "Stemming in the Language Modeling Framework," *SIGIR*, *ACM*, Toronto, 28 July-1 August 2003.

[14] A. Farag and N. Andreas, "N-Gram Conflation Approach for Arabic Text," *SIGIR*, *ACM*, Amsterdam, 7 July 2007.

[15] K. Dinesh and R. Prince, "Stemming of Punjabi Words by Using Brute Force Technique," *International Journal of Engineering Science and Technology*, *IJEST*, Vol. 3, No. 2, 2011.

[16] D. Sajib and N. Vincent, "Unsupervised Morphological Parsing of Bengali," Lang Resource & Evaluation, Springer, 2007.

[17] R. Monica, M. Scott and Y. Yiming, "Unsuperised Learning of Arabic Stemming Using a Parallel Corpus," *Proceeding of the* 41*st Annual Meeting of the Association for Computation Linguistics*, July 2003, pp. 301-398.

[18] N. S. Giridhar, K. V. Prema and N. V. Subba Reddy, "A Prospective Study of Stemming Algorithms for Web Text Mining," *Ganapt University Journal of Engineering & Technology*, Vol. 1, 2011, pp. 28-34.

[19] K. Chouvalit and B. Veera, "Inverted Lists String Matching Algorithms," *International Journal of Computer Theory and Engineering*, Vol. 2, No. 3, 2010, pp. 352-357.

[20] K. Koudas, S. Sunita and S. Divesh, "Record Linkage: Similarity Measures and Algorithms."

[21] J. Ms. Anjali, "A Comparative Study of Stemming Algorithms," *IJCTA*, Vol. 2, No. 6, 2011, pp. 1930-1938.

[22] B. Bal Krishna and S. Prajol, "A Morphological Analyzer and a Stemmer for Nepali," Madan Puraskar Pustakalaya, Working Papers 2004-2007.

[23] F. Cuna Ekmekcioglu, L. F. Michael and W. Peter, "Stemming and N-Gram Matching for Term Conflation in Turkish Texts," *Information Research*, Vol. 2, 1996.

[24] C. Sitaula, "Semantic Text Clustering Using Enhanced Vector Space Model Using Nepali Language," *GESJ*, Vol. 36, No. 4, 2012, pp. 41-46.