Modelagem de Jogos Massivos On-Line como um Processo de Salto

Izabella Araújo Rodrigues Alves Santos

Dissertação apresentada ao
Programa de Pós-Graduação em Estatística da
Universidade Federal de Minas Gerais
para obtenção do título de Mestre em Estatística

Orientadora: Profa. Dra. Denise Duarte (UFMG)
Co-orientador: Prof. Dr. Marcelo Azevedo (UFMG)

Belo Horizonte
6 de outubro de 2013

Esta versão da dissertação contém as correções e alterações
sugeridas pela banca durante a defesa da versão original
do trabalho, realizada em 05 de setembro de 2013.

Comissão Julgadora:

- Profa. Dra. Denise Duarte (orientadora) - UFMG
- Prof. Dr. Marcos Prates- UFMG
- Prof. Dr. Fábio Machado - USP

Agradecimentos

A Deus dedico este trabalho, "Pois o Senhor é quem dá sabedoria; de sua boca procedem o conhecimento e o discernimento."Provérbios 2:6. Àquele que mudou a minha história, me capacitou e me guiou até aqui seja dada toda honra e toda glória. Porque ele me amou e me tirou de um caminho de tristezas e incertezas e me fez alcançar muito mais do que meus medos e complexos me permitiam sonhar. Obrigada Senhor.

Obrigada mãe por ser uma alegria e um descanso constantes. Por se preocupar tanto comigo e se oferecer, tantas vezes, para me ajudar com os trabalhos e estudos mesmo sem nem saber o que é desvio padrão. Agradeço também ao meu pai, que apesar de não estar vivo para acompanhar esta minha conquista, sempre foi e será um exemplo de empenho, dedicação, superação e honestidade. Obrigada por me mostrar com sua própria história que não importa de onde se veio mas sim onde se quer chegar. Obrigada Matheus, meu amor, por me consolar nas horas de desespero, por se alegrar com minhas conquistas e pela imensa ajuda com a dissertação, desde a primeira ideia até as conclusões finais. Sem sua ajuda este trabalho não teria sido feito.

Obrigada Denise por dividir comigo um pouco do seu imenso conhecimento e suas ideias, por ser tão paciente, presente e dedicada, por se preocupar não só com meu trabalho mas também comigo. Obrigado por ser tão humana. Obrigada Marcelo pela enorme boa vontade, sugestões e ideias.

Obrigada a todos os amigos que fiz durante o mestrado, que me ajudaram com os trabalhos e dúvidas incessantes, que me fizeram rir em muitas das vezes em que a vontade era chorar. Obrigada especialmente à Sil, Nívea, Talita, Mariana, Gabriel e Paulo. Como vocês foi bem mais divertido.

Obrigada aos professores Marcos Prates (UFMG) e Fábio Machado (USP), membros da banca, pelas correções e sugestões que ajudaram a melhorar este trabalho.

Por fim, obrigada ao CNPq pelo apoio financeiro durante a execução desse trabalho, e ao Departamento de Estatística da UFMG pela estrutura e suporte oferecidos.

"É preciso que o discípulo da sabedoria tenha o coração grande e corajoso. O fardo é pesado e a viagem longa."

(Confúcio)

Introdução

A indústria de jogos virtuais tem crescido nos últimos anos. Em um estudo realizado pelo Programa infoDev fundado pelo Departamento para Desenvolvimento Internacional (DFID) do Reino Unido, Lehdonvirta e Ernkvist (2011) mostram que moedas virtuais on-line e de trabalho digital têm oferecido oportunidades reais de renda para trabalhadores menos qualificados dos países em desenvolvimento. Eles descobriram que mais de 100.000 pessoas em países como China e Índia ganham a vida através jogos on-line e sites. Eles estimaram que o mercado para esses serviços de jogos de aluguel valia cerca de US$ 3 bilhões em 2009. Segundo os autores, World of Warcraft (WoW), publicado pela Blizzard Entertainment, é atualmente o principal jogo on-line com mais de 12 milhões de jogadores ativos em 2010. World of Warcraft é um jogo on-line massivo, popularmente conhecido como MMO (Massively Multiplayer On-line). Um MMO é um tipo de jogo que permite que milhares de usuários criem personagens em um mundo virtual dinâmico e joguem simultaneamente através Internet.

Com a popularidade dos jogos virtuais, surgiu o desafio de manter milhares de personagens interagindo no mesmo mundo virtual de modo a que a qualidade da experiência do usuário não seja comprometida. De acordo com Santos (2010), nos servidores destes jogos ocorre grande tráfego de uma massa heterogênea de pequenas tarefas em um curto período de tempo. Isso, juntamente com outras características, gera um desafio na manutenção do bom desempenho desses jogos. O estudo do desempenho destes sistemas cria oportunidades de melhoria no jogo a fim de contribuir para a qualidade da experiência do usuário e reduzir os custos de infra-estrutura. Realizar este tipo de estudo com os servidores do jogo é invasivo, tem um grande impacto sobre a operação e requer o empenho e colaboração dos mantenedores do sistema. Uma abordagem comumente utilizada para estudar estes sistemas complexos, sem a utilização do sistema real, é a simulação. Para implementar um simulador capaz de reproduzir o comportamento da carga gerada pelos personagens de um jogo on-line no servidor é necessário implementar de um modelo capaz de descrever o movimento dos personagens no jogo. A simulação deve representar de forma confiável a movimentação de personagens no jogo usando um modelo adequado. Neste caso, os parâmetros do modelo devem ser estimados usando dados do jogo.

Devido à dificuldade em coletar dados de jogos on-line, o que pode desestabilizar o sistema, pouco tem sido estudado sobre a movimentação dos personagens de jogos on-line.

Existem alguns estudos que analisam outras características destes jogos, como tráfego de usuários, consumo de banda, tamanho do pacotes e tempo de chegada de pacotes (Chen et al, 2006 ; . Svoboda et al, 2007 ; . Fang e Wenli , 2006). La e Michiardi (2008) realizaram uma análise dos dados de mobilidade dos personagens no jogo Second Life , um mundo virtual on-line, utilizando redes espaciais e temporais eles caracterizam a distribuição estatística de oportunidades de contato entre os usuários como uma lei de potência truncada e encontram semelhanças entre a distribuição do tempo de contato em ambientes virtuais e aqueles obtidos em experimentos reais. Santos (2010) comparara a mobilidade dos personagens do WoW com Second Life e com o modelo de mobilidade aleatória. Ele identifica algumas características que tornam estes ambientes semelhantes uns aos outros e ao comportamento humano no mundo real. Tan et al (2005) apresentou um modelo de mobilidade de carga de trabalho , Networked Game Mobility Model (NGMM), para a representar mobilidade em um jogo Atirador em Primeira Pessoa (First Person Shoter - FPS), que não é um MMO. O modelo é uma extensão do Caminho de Ponto Aleatório (Random Way Point - RWP) que consiste em dois estados: fixo e móvel. Quando o destino é atingido, ele faz uma pausa por um período de tempo antes de selecionar um novo destino. Eles modelam destino, velocidade, direção e tempo de pausa. O modelo NGMM é aplicável apenas a jogos FPS. Miller e Crowcroft (2009) analisaram algumas medidas dos movimentos de caracteres em campos de batalha de WoW. Eles constataram que a maioria dos movimentos dos personagens entre os objetivos é individual e não coletivo.

Neste trabalho, propomos um modelo estatístico para descrever a mobilidade dos personagens no mapa de um jogo MMO, assim como as estimativas para este modelo. Apresentamos também um simulador para a mobilidade. Analisamos os dados geográficos de posição dos personagens no mapa do jogo WoW coletados por Santos (2010). No WoW, os jogadores assumem o papel de personagens em um ambiente fictício povoado por duas facções inimigas: Aliança e Horda. O mapa do jogo é dividido em várias regiões geograficamente delimitadas, chamadas zonas. O principal objetivo do jogo é avançar através dos níveis. Os personagens começam no nível um e aumentam o seu nível matando monstros e realizando missões. Algumas missões exigem viajar para outras zonas, logo, ocorrem um grande número de transições entre zonas com missões comuns. Cada zona tem monstros de uma determinada faixa de níveis, zonas geograficamente próximas, normalmente, têm monstros de níveis consecutivos. Em geral, os personagens ficam em uma zona com monstros de níveis próximos ao seu, até atingir um certo nível que lhes permite avançar para outra zona em que os monstros têm um nível mais avançado e assim por diante. No decorrer do jogo, os personagens precisarão de suprimentos, tais como armaduras, armas, comida, entre outros. Estes suprimentos podem ser comprados em zonas em que existem centros comerciais. Assim, os personagens tendem a mover-se para as zona com centro comercial mais próxima regularmente. Portanto, acreditamos que a posição atual de um personagem no mapa do jogo carrega informação preditiva sobre a sua posição futura. Em outras palavras, podemos prever probabilisticamente a próxima zona para a qual ele irá se mover, com base em sua posição atual. Por esta razão, propomos um modelo estocástico para a mobilidade no jogo. Uma vez que o tempo de permanência em uma região do mapa é uma variável aleatória contínua, modelamos a mobilidade como um processo de salto. Nós definimos os estados do processo como regiões do mapa (zonas) e modelamos a movimentação entre as regiões de mapa como transições entre os estados do processo. Portanto, o personagem irá se mover no mapa de jogo de acordo com as transições do modelo. Utilizando a metodologia proposta, obtivemos um modelo para descrever o movimento dos personagens no jogo. Propusemos duas diferentes abordagens para modelar o tempo de permanência em cada estado do processo: paramétrica e não paramétrica. Aplicamos o Teste de Kolmogorov-Smirnov para mostrar que os dados simulados através da metodologia proposta eram consistentes com os dados observados.

Este artigo está organizado da seguinte forma. A Seção 2 apresenta uma breve descrição do dados. Na Seção 3 apresentamos a metodologia de um processo de salto, o método utilizado para determinar os estados do processo e a metodologia do estimador de densidade kernel. A descrição do simulador, o algoritmo de simulação e um exemplo de conjunto de dados obtidos no final da simulação são apresentados na Seção 4. As Seções 5 e 6 apresentam os resultados e as conclusões, respectivamente.

Modeling Mobility in Massively Multiplayer On-line Games as a jump process.

Izabella A. R. A. Santos[a], Denise Duarte[a], Marcelo Azevedo Costa[b]

[a] *Departamento de Estatística, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Sala 4060, Bairro Pampulha, Belo Horizonte, Minas Gerais - Brasil*

[b] *Departamento de Engenharia de Produção, Escola de Engenharia, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Sala 3108, Bairro Pampulha, Belo Horizonte, Minas Gerais - Brasil*

Abstract

This paper addresses a methodology to model the mobility of the characters in Massively Multiplayer On-line (MMO) Games. Despite the importance of the virtual games industry, most studies in this area are superficial and merely descriptives. We propose to model the mobility of characters in the map of an MMO game as a jump process using two approaches to model the times spent in the states of the process: Parametric and Non-Parametric. Furthermore, a simulator for the mobility is presented. We analyze geographic position data of the characters in the map of the game World of Warcraft (WoW) and compare the observed and simulated data.

Keywords: Jump process, on-line games, MMO, character mobility.

1. Introduction

The virtual games industry has grow in recent years. In a study by the World Bank Group's infoDev program, funded by the United Kingdom's Department for International Development (DFID), Lehdonvirta and Ernkvist (2011) claim that virtual on-line currencies and digital work now provide real income opportunities for poor and unskilled workers in developing countries. They found that more than 100,000 people in countries such as China and India earn a living through on-line games and websites. They estimated the market for such gaming-for-hire services to be worth around \$3 billion in 2009. World of Warcraft (WoW), published by

Email addresses: `izabellaa@est.mest.ufmg.br` (Izabella A. R. A. Santos), `dduarte.est@gmail.com` (Denise Duarte), `macosta.est@gmail.com` (Marcelo Azevedo Costa)

Blizzard Entertainment, is currently the leading global on-line game with over 12 million active player accounts as of 2010 (Lehdonvirta and Ernkvist (2011)). The World of Warcraft is a massive on-line game, popularly known as MMO (Massively Multiplayer On-line). An MMO is a type of game that allows thousands of users to create characters in a dynamic virtual world and play simultaneously through Internet.

With the popularity of virtual games came the challenge of maintaining thousands of characters; interacting in the same virtual world so that the quality of the user experience is not compromised. According to Santos (2010), the servers of these games experience large mass traffic of heterogeneous small tasks in a short period of time. This, together with other characteristics, generates a challenge in maintaining good performance of those games. The study of the performance of those systems creates opportunities for improvement in the game in order to contribute to the quality of the user experience and reduces infrastructure costs. Performing this type of study using the game servers is invasive and has a great impact on the operation. It requires the commitment and cooperation of the maintainers of the system. One commonly used approach to study such complex systems, without the use of the real system, is simulation. To implement a simulator able to reproduce the behavior of the load which the characters in a game on-line generate on a server is necessary to implement a model capable of describing the movement of characters in the game. The simulation must represent reliably the movement of characters in the game using a proper model. In this case the parameters of the model must be estimated using data from the game.

According Santos (2010), using data simulated from virtual worlds may be possible to study some real world events as the spread of diseases or computer virus, once there are similarities between the mobilities in these two worlds and, it is easier extracting the position data of users in virtual worlds than collect data from people in the real world. Furthermore, may be also possible to study, for example, network performance delay tolerant (DTNs), mesh networks (mesh) and problems in wireless networks related to mobility (eg, Bluetooth and WiFi).

Due to the expense of providing on-line game data, which may crash the system, little studied about the movement of the characters in on-line games has been done. There are some studies analyzing other features of these games, such as player traffic, band consumption, packet size, time of packet arrival and number of users (Chen et al., 2006; Svoboda et al., 2007; Fang and Wenli, 2006). La and Michiardi (2008) performed an analysis of mobility data of the characters in Second Life, an on-line virtual world, using metric spatial, and temporal, and complex networks. They characterized the statistical distribution of contact opportunities among users as a power-law with cutoff distribution and found similarities between the contact-time distributions in virtual environments and those obtained in real-world experiments. Santos (2010) compared the mobility of the characters from WoW with Second Life and with random mobility model. He identified some features which make these environments similar to each other and similar to human behavior in the real world. Tan et al. (2005) presented a workload mobility model, Networked Game Mobility Model (NGMM), for synthesizing mobility in a First-Person-Shooter (FPS) game which is not an MMO. Their model is an extension of Random Way Point (RWP) mobility which consists of two states: stationary and moving. When the destination is reached, it pauses for a period of time before selecting a new destination. They modeled destination, speed, direction and pause time. The NGMM model is only applicable to FPS games. Miller and Crowcroft (2009) analyzed a few measures of character movements in WoW battlegrounds. They found that the majority of character movements between objectives is individual, not collective.

In this paper, we propose a statistical model to describe the mobility of characters in the map of an MMO game, and the estimates for this model. We also present a simulator for the mobility. We analyze the geographic position data of the characters in the WoW game map collected by Santos (2010). In the game, the players assume the roles of characters in a fictional setting populated by two enemy factions: Alliance and Horde. The game map is divided into geographically delimited regions, called zones. The main goal of the game is to advance through the levels. The characters start at level one and increase their level by killing monsters and performing quests. Some quests require travel to other zones. Therefore a large number of transitions between zones with common quests will occur. Each zone has monsters of a certain range of levels, and geographically proximate zones usually have monsters of consecutive levels. Usually, the characters will stay in a zone with monsters having levels close to theirs until attaining a certain level that allows them to advance to another zone in which the monsters also have a more advanced level and so on. Through the game, the characters will need supplies such as armor, weapons, food, among others. Supplies can be bought in zones in which there are merchant centers. Therefore the characters tend to move to the closest merchant center zones, regularly. We believe that the current position of a character in the game map carries predictive information about his future position. In other words, we can predict probabilistically the next zone that he will go to, based in his current position. For this reason, we propose a stochastic model for mobility in the game. Since the time spent in a map region is a continuous random variable, we model the time as a jump process. We define the states of the process as map regions (zones) and we model the movements between map regions as transitions between the states of the process. Then, the character will move in the game map according to the transition model.

Using the proposed methodology, we found a model to describe the movement of characters in the game. We propose two different approaches to model the time spent in each state of the process: parametric and non-parametric. We applied the Kolmogorov-Smirnov test to show that the data simulated through the proposed methodology are consistent with the observed data.

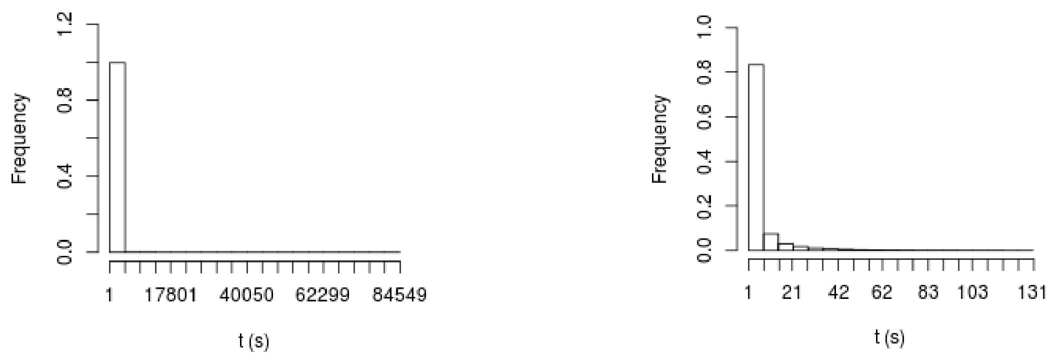
This paper is organized as follows. Section 2 shows a brief description of the data. In Section 3 we present the methodology of a jump process, the method used to determine the states of the process, and the methodology of the kernel density estimator. In Section 4 we give a description of the simulator, the simulation algorithm and an example of the data set obtained at the end of the simulation. The Sections 5 and 6 present the results and conclusions, respectively.

2. The Data

The data were collected from a Brazilian private server of the WoW game for 24 hours, during a regular weekday. At each 0.3 consecutive seconds (approximately) the positions of the characters in the virtual Cartesian coordinate system and the time of the day were recorded. Posteriorly the times spent in each geographical position were computed. Therefore the data set was compound of four columns, the identifier of the character, the geographical coordinates and the time that the character spent in each position.

The WoW map consists of three continents: Eastern Kingdoms, Kalimdor and Northrend. In our study, we considered only the Eastern Kingdoms continent, which is the most popular region between the players. Currently, there are four expansions of the game. The server in which the data were collected, worked with the Burning Crusade expansion. In this version, the Eastern Kingdom continent is divided in 27 zones. We obtained data from 22 of the 27 zones.

Because of the great impact caused by the data acquisition, the server went off-line some times. This is one of the reasons why is so difficult to have the cooperation of the maintainers of those games for collect data. There was no records of the times when the server went off-line or the time it stayed off-line. When the server went on-line again the characters returned to the last position registered and the time while the server was off-line was added to the time that the character spent in the position where he was when the server went off-line. We observed a few values of times in the data set which were much higher than the most times. We believe that those times are the times when the server was off-line. Therefore we found the quantile 99% of the times and eliminated the times higher than it. Figures 1a and 1b shows the histograms of the times before and after the removal of the extreme times.



(a) Before. (b) After.

Figure 1: Histograms of the times before and after we remove the extreme values of times.

The number of observations collected in the Eastern Kingdoms was 250,417, the maximum time observed was 83,710s and the 99% quantile was 129.36s. Table 1 shows some descriptives statistics of the times spent in the map positions after the removal of the values higher than 129.36s. More details about the method of data collection and a descriptive analysis of the data can be found at Santos (2010).

Table 1: Descriptives statistics of the times spent in the map positions - Times in seconds.

N	Min	Q1	Median	Mean	Q3	Max
247911	1.02	1.58	2.67	6.352	5.36	129.3

In this paper we will use the following terminology:

- Character: Avatar in the virtual world controlled by a player in the real world;
- Zone: Region in the game map defined by the game publisher. These regions have borders like a state or country;
- Session time: Time in which the character remained logged.

3. Model

Given the current zone of the character, it is expected that the next zone is the closest merchant center or a zone with monsters of higher level or a zone related to its by quest. In all three cases, the zone change will occur after a certain time. When the character needs supplies it goes to a merchant center, buys what it needs, and returns to its previous zone or goes to another one. Eventually, it will need new supplies. When the character moves to a zone with higher level monsters, the next time when he will look for a zone with monsters of a higher level will occur after he reach a higher level, what will happen after a certain time. The quests also take time to be performed. If he moves to other zone because of a quest, he will spend some time performing the quest before moving again. Therefore we are working with a temporal process, the zone transitions occurs over time which led us to model the mobility of the characters as a stochastic process. In this case, the time is a continuous random variable, then we modeled it as a jump process. The following will be presented the methodology of a jump process.

3.1. Jump Processes

Let $\mathbf{\Gamma}$ be the set of all states of a dynamical system and $x, y \in \mathbf{\Gamma}$ the states of the system. A jump process is defined by a random variable $X(t)$, $t \in [0, \infty)$, which starts in a state x_0 at time $t = 0$ and stays in x_0 until some time t_1 when the process jumps to another state x_1. The process stays in x_1 until reaching some time $t_2 > t_1$ in which it jumps to another state x_2 and so forth. If the process is in a state y, it will make a transition to another state x according to $r(y, x)$, such that $r(y, y) = 0$ and $\sum_x r(y, x) = 1$. Once the process is in x, the time spent in this state is a random variable which follows some distribution function $F_x(t)$.

Let τ be the random variable time spent in a state. Each state can have a different distribution $F_x(t)$. The time spent in a state y and the choice of the next state x are assumed independent random variables, so then

$$P(\tau \leq t, X(\tau) = x, X(0) = y) = r(y, x)F_y(t). \tag{1}$$

The jump process is called Markov jump process if, and only if, $F_y(t)$ is the Exponential distribution. In this case, the process presents the Markov property:

Given times $0 < t_1, t_2 < ... < t_n < s$ and $t > 0$,

$$P(X(t+s) = x | X(s) = y, X(t_n) = x_n, ..., X(t_1) = x_1) = P(X(t+s) = x | X(s) = y). \tag{2}$$

According to Equation 2, given a set of previous states at earlier times, the Markov jump process forgets all but the state at the most recent time. In this case, Equation 1 gives the conditional transition probability that the jump process is in state x at time t given that it was in state y at any time, and not only in time 0. In other words, the jump process starts all over again at this most recent time. For more details about jump process see Ferrari and Galves (2000) and Feng (2004).

3.2. Clustering

The first step to implement the jump process model is to find the state space since it is unknown. The logical choice seemed define it as the zone set of game map. Unfortunately the boundaries of the game zones were not available, then we used the clustering k-means method, see Hartigan and Wong (1979), with the clusters number equal to the number of map zones (22) and the clusters centers given by the geographic centers of map zones. The figures 2a and 2b shows the Eastern Kingdoms map and the clustering result, respectively. Nearby points with the same color belong to the same cluster. Visually, the boundaries K-means boundaries were very close to the boundaries of the map zones.

(a) Eastern Kingdoms map.

(b) Clustering results.

Figure 2: Eastern Kingdoms map and clustering result over the map.

During the analysis, we observed that the time spent in one of the 23 states behaved different from others. We found that this state was the Duskwood zone. In the middle of this zone, there is a region called Twilight Grove. Twilight Grove is home to a portal to a region outside the boundaries of the physical world of the game, in other words, this region does not exists in the game map, and we did not have data in this area. Therefore we decided to treat Duskwood as two states, one at the east side of the Twilight Grove and other at the west side. The Figures 3a and 3b shows the map of Duskwood and the division of the data of this zone in two new states. Therefore the process went on to have 23 states. After clustering we obtained a total of 2667 observations of times spent in the map zones, in which the larger number of observations per zone was 760 and the minimum was 21.

(a) Duskwood map. (b) Division of the Duskwood data.

Figure 3: Duskwood map and division of the Duskwood data in two states.

3.3. Estimation

Let $n_{y,x}$ be the number of transitions from state y to x. Using the Equation 1 we can write the likelihood of the process as:

$$L(y,x;t) = \prod_{x\in\mathbf{\Gamma}} \prod_{\substack{y\in\mathbf{\Gamma},\\ y\neq x}} \left[r(y,x)F_y(t)\right]^{n_{y,x}},$$

then taking the logarithm we have:

$$l(y,x;t) = \sum_{x\in\mathbf{\Gamma}} \sum_{\substack{y\in\mathbf{\Gamma},\\ y\neq x}} n_{y,x}\left[log(r(y,x)) + log(F_y(t))\right],$$

$$l(y,x;t) = \sum_{x\in\mathbf{\Gamma}} \sum_{\substack{y\in\mathbf{\Gamma},\\ y\neq x}} n_{y,x} log(r(y,x)) + \sum_{x\in\mathbf{\Gamma}} \sum_{\substack{y\in\mathbf{\Gamma},\\ y\neq x}} n_{y,x} log(F_y(t)). \tag{3}$$

We can see that the estimation by maximum likelihood can be done in two independent parts, discrete, the first part of the Equation 3, and continues, the second part.

For reasons of computational cost, and for being a reasonable choice, we consider that the process which governs the character movement changes its state by an order one process, then the transition probabilities were estimated by the maximum likelihood estimator,

$$\hat{r}(y, x) = \frac{n_{y,x}}{n_y},$$

where n_y is the number of times when the process was in y, then knowing the distribution that governs the times spend in the states of the process, Equation 1 can be estimated by:

$$\hat{P}(\tau \leq t, X(\tau) = x, X(0) = y) = \frac{n_{y,x}}{n_y} F_y(t),$$

3.3.1. Non Parametric Estimation

If it is not possible to determine the distribution of the times spent in the states of the process, it can be estimated by kernel density estimation. Let $Y_1, Y_2, ..., Y_n$ be a sample of size n from a random variable with density f. The kernel density estimator of f at the point y is given by

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right), \tag{4}$$

where the kernel K is a function satisfying $\int K(y)dy = 1$ and h is the smoothing parameter known as bandwidth. The function K is generally chosen to be a unimodal probability density symmetric about zero, a common choice is the Gaussian kernel,

$$K(y) = \frac{1}{\sqrt{2\pi}} exp - \frac{y^2}{2}.$$

There are many rules to choose the bandwidth, in this paper we will use the method of cross-validation proposed by Hall (1983) Hall, P. (1983), the direct plug-in approach proposed by Sheather and Jones (1991) and the method proposed by Silverman (1986), known as *rule of thumb*. Basically, the three methods consist of minimizing the asymptotic mean squared error (AMISE) of the kernel estimator for the density. The difference between the methods is the form of estimating the part of the AMISE which depends on the density function of the data which is unknown. Therefore we can estimate the equation that characterizes the movement process in the game map, by the nonparametric approach, using Equation 1 with the kernel density estimator, the Gaussian kernel and integrating Equation 4 from 0 to t. Then we find:

$$\hat{P}(\tau \leq t, X(\tau) = x, X(0) = y) = \left(\frac{n_{y,x}}{n_y}\right) \frac{1}{n_y} \sum_{i=1}^{n_y} \left[\Phi\left(\frac{t - Y_i}{h_y}\right) - 0.5\right],$$

where h_y is the bandwidth in the state y, n_y is the number of times that the process was in y and Φ is the cumulative distribution function of the standard normal distribution.

4. Simulator

The simulator was implemented in R software language. Despite that R does not has a very good performance compared to others languages, the software is free and its language is easy to learn and to extend with functions written by the user. Also, most of the statistics methods used in this paper were already implemented in R. The simulator was built in way that the user must specify the number of characters, the minimal time spent in a state, the session time which can be different for each character, and the method of density estimation for the time spend in the process:parametric or nonparametric. subprocess. If the user choses the parametric method, he needs to specify the distribution, Weibull, Gamma or Exponential. He can specifies the distribution parameters or let them be estimated by the maximum likelihood estimator using a sample data. If he choses the nonparametric method, he needs to specify the rule to choose the bandwidth. The options are those implemented in the density function of the R software, Sheather and Jones (1991), Scott (1992), Silverman (1986), biased and unbiased cross-validation (Sheather (2004)). The simulation was executed in a notebook computer with a Intel core i5-3317U CPU @ 1.70GHz and 4 GB of RAM.

Algorithm:

1. Estimate the transitions matrix between states from the sample (also can be pre-specified by the user).
2. Estimate the probability of the character starts the game in each state.
3. Generate a data set containing the initial states of the characters and the times they will remain in these states.
4. Generate new observations for each character with the states to which they moved, and the new times spent in those states.
5. Repeat the previous step until the sum of the times of each character is equal to the total time of game specified.

At the end of the simulation, the user will obtain a data set with three columns: the identifier of the character, the state in which he is and the time spent in this state. Figure 4 shows an example of the first ten observations of the data set obtained at the end of a simulation with five characters, 12 hours of game for each one and time simulated by the Weibull distribution.

R Data Editor

Copy Paste Quit

	ID	State	t
1	1	20	352.343
2	2	14	317.841
3	3	14	307.665
4	4	14	1375.523
5	5	19	365.085
6	1	14	267.992
7	2	23	6.997
8	3	20	555.289
9	4	19	611.098
10	5	14	259.948

Figure 4: Example of the data set obtained at the end of the simulation.

The simulator code and help material can be downloaded from http://www.est.ufmg.br/ftp/denise/MOGMS/.

5. Results and Discussion

Once we had defined the states space of the process we need to estimate the distribution that governs the times spent in these states. Table 2 shows some descriptives measures of the times spent in each state.

The most visited zone during the data collection was Elwynn Forest (14). It is the most popular zone between the characters of the Aliance faction, which is the most popular faction. Elwynn Forest has a large commercial center located at the Stormwind castle. In mean, it is the third zone where the characters spent more time. The first one is Stranglethorn Vale (19). Stranglethorn Vale also has a large commercial center located at the Booty Bay city. It is a neutral city, in other words, in this city the characters of the two enemy factions coexist peacefully. The zone less visited was Searing Gorge (21). For the Alliance characters, the entrance to this zone is locked until the key is obtained via a quest. Once not all characters complete the quests, this is probably the reason for the low number of visitations.

Table 2: Descriptives measures of the times (seconds) spent in the process states.

State	Zone	N	Min	Q1	Q2	Mean	Q3	Max
1	Alterac Mountains	38	1.75	30.86	128.8	488	512	2949
2	The Hinterlands	33	1.02	8.05	98.2	410	263	5941
3	Arathi Highlands	29	4.38	18.98	161.1	336	401	2398
4	Badlands	23	2.58	21.26	134.1	355	554	1808
5	Blasted Lands	141	1.22	54.92	134.1	342	445	3567
6	Burning Steppes	50	3.03	64.62	407.9	998	1307	8967
7	Wetlands	37	5.91	39.13	235.0	653	935	4589
8	Deadwind Pass	53	1.14	36.69	117.9	203	230	1879
9	Dun Morogh	174	1.03	33.64	108.5	390	356	4727
10	Westfall	84	1.73	82.49	285.2	735	1083	6542
11	Duskwood1	190	1.02	29.02	127.3	383	314	5625
12	Western Plaguelands	27	3.25	37.57	146.2	333	300	1594
13	Eastern Plaguelands	26	3.78	87.62	219.1	511	474	3600
14	Elwynn Forest	760	1.08	111.90	352.4	783	923	8914
15	Tirisfal Glades	151	1.95	98.77	255.9	654	697	6242
16	Hillsbrad Foothills	82	1.08	31.13	113.9	514	397	8009
17	Swamp of Sorrows	156	1.12	35.42	95.0	239	223	2382
18	Loch Modan	33	1.09	8.09	61.4	361	254	4677
19	Stranglethorn Vale	261	1.14	133.90	414.7	1002	1263	11410
20	Redridge Mountains	78	1.44	71.30	241.9	510	759	4494
21	Searing Gorge	13	2.67	24.03	79.4	427	893	2288
22	Silverpine Forest	40	6.17	41.58	188.0	399	565	1871
23	Duskwood2	188	1.02	20.69	85.6	314	303	3106

To evaluate the goodness of fit of the proposed models for the time spent in the states of the process we plotted the QQ-plots and computed the Kolmogorov-Smirnov test (Conover (1971)). The Kolmogorov-Smirnov statistic quantifies the distance between the distribution functions of two samples or one sample and a reference probability distribution. As the process has a high number of states, and all of them showed a similar behavior, we will present in this paper the QQ-plots of only a few selected states.

The figures from A.5 to A.7 shows the QQ-plots with the 95% confidence limits for the fit of the parametric models in three of the 23 the states of the process and the Table 3 shows the Kolmogorov-Smirnov test results in all states. The null hypotheses of the test is that the sample times are drawn from the proposed distribution. In parentheses are the p-values of the tests. We will reject the null hypothesis, if the p-value of the test is less than 0.05.

Table 3: Kolmogorv-Smirnov test for the parametric models.

State	Zone	Exponential	Gamma	Pareto	Log-Normal	Weibull
1	Alterac Mountains	0.323(**0.000**)	0.141(0.396)	0.974(**0.000**)	0.103(0.774)	0.101(0.799)
2	The Hinterlands	0.207(**0.000**)	0.091(**0.026**)	0.996(**0.000**)	0.053(0.466)	0.072(0.130)
3	Arathi Highlands	0.337(**0.001**)	0.147(0.433)	0.970(**0.000**)	0.101(0.858)	0.088(0.941)
4	Badlands	0.246(0.050)	0.129(0.673)	0.966(**0.000**)	0.144(0.540)	0.128(0.685)
5	Blasted Lands	0.327(**0.011**)	0.152(0.608)	0.957(**0.000**)	0.134(0.751)	0.132(0.768)
6	Burning Steppes	0.191(**0.000**)	0.104(0.093)	0.986(**0.000**)	0.056(0.761)	0.076(0.397)
7	Wetlands	0.298(**0.000**)	0.128(**0.004**)	0.988(**0.000**)	0.055(0.631)	0.083(0.148)
8	Deadwind Pass	0.288(**0.000**)	0.109(0.551)	0.980(**0.000**)	0.140(0.257)	0.102(0.642)
9	Dun Morogh	0.273(**0.006**)	0.114(0.676)	0.973(**0.000**)	0.095(0.864)	0.104(0.776)
10	Westfall	0.170(0.176)	0.080(0.944)	0.975(**0.000**)	0.105(0.726)	0.080(0.941)
11	Duskwood1	0.117(0.432)	0.109(0.520)	0.981(**0.000**)	0.193(**0.033**)	0.130(0.308)
12	Western Plaguelands	0.445(**0.007**)	0.219(0.495)	0.923(**0.000**)	0.176(0.753)	0.177(0.749)
13	Eastern Plaguelands	0.224(**0.000**)	0.091(0.467)	0.988(**0.000**)	0.066(0.836)	0.073(0.733)
14	Elwynn Forest	0.269(**0.000**)	0.134(**0.002**)	0.984(**0.000**)	0.098(0.051)	0.081(0.168)
15	Tirisfal Glades	0.148(0.058)	0.078(0.699)	0.987(**0.000**)	0.141(0.082)	0.095(0.456)
16	Hillsbrad Foothills	0.222(0.120)	0.150(0.530)	0.963(**0.000**)	0.098(0.934)	0.117(0.814)
17	Swamp of Sorrows	0.372(**0.000**)	0.151(0.403)	0.969(**0.000**)	0.101(0.855)	0.095(0.896)
18	Loch Modan	0.215(0.156)	0.129(0.729)	0.962(**0.000**)	0.110(0.881)	0.116(0.837)
19	Stranglethorn Vale	0.215(**0.000**)	0.135(**0.007**)	0.994(**0.000**)	0.062(0.586)	0.100(0.090)
20	blackridge Mountains	0.158(**0.000**)	0.057(**0.014**)	0.999(**0.000**)	0.054(**0.025**)	0.035(0.311)
21	Searing Gorge	0.187(**0.000**)	0.104(0.077)	0.993(**0.000**)	0.081(0.269)	0.072(0.420)
22	Silverpine Forest	0.339(**0.000**)	0.146(0.055)	0.988(**0.000**)	0.057(0.940)	0.074(0.729)
23	Duskwood2	0.281(**0.000**)	0.129(**0.006**)	0.994(**0.000**)	0.032(0.995)	0.087(0.145)

As we can see, the sample times did not fit to the Pareto distribution in any of the states. The exponential, gamma and log-normal distributions did not fitted to all states. The Weibull distribution was the only one capable of fit to spent times in all states. Therefore we used the Weibull distribution in the simulations. The parameterization of the Weibull distribution used was the following:

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0,$$

where $k > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. To estimate the parameters, we used the maximum likelihood estimators. Differentiating the logarithm of the likelihood function with respect to k and λ, and subsequently eliminating λ, the following equation can be found:

$$\frac{\sum_{i=1}^{n} x_i^k ln(x_i)}{\sum_{i=1}^{n} x_i^k} - \frac{1}{k} - \frac{1}{n}\sum_{i=1}^{n} ln(x_i) = 0. \tag{5}$$

The MLE of k can be found by solving the Equation 5 using a numerical procedure as Newton-Raphson. Then the MLE of λ can be calculated by:

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

Using the Weibull distribution, Equation 1 can be written as:

$$\hat{P}(\tau \leq t, X(\tau) = x, X(0) = y) = \left(\frac{n_{y,x}}{n_y}\right)\left(1 - e^{-(t/\lambda_y)^{k_y}}\right),$$

where k_y and λ_y are the Weibull parameters in the state y.

Once the times did not follow the Exponential distribution, they do not present the Markov property. Therefore the conditional probability that the process will be in the state y at time t given the states at previous times can be dependent on all the past states, not just the most recent as in the Markov jump process. In this case, Equation 2 does not apply to the process.

sTable 4 shows the values of the parameters in each state and the p-value of the Kolmogorov-Smirnov test for the hypothesis that the times follow the Weibull distribution with the parameters specified in the table.

Furthermore, we tried to find a relationship between the values of the parameters and some characteristics of the states, like level, existence of commercial center, size

Table 4: Parameters of the Weibull in the states process.

States	Zones	k	lambda
1	Alterac Mountains	0.567	302
2	Arathi Highlands	0.497	181
3	Badlands	0.664	250
4	Blasted Lands	0.600	242
5	Burning Steppes	0.734	277
6	Deadwind Pass	0.594	664
7	Dun Morogh	0.603	434
8	Duskwood1	0.771	172
9	Duskwood2	0.628	261
10	Eastern Plaguelands	0.680	566
11	Elwynn Forest	0.593	235
12	Hillsbrad Foothills	0.680	252
13	Loch Modan	0.736	412
14	Redridge Mountains	0.724	628
15	Searing Gorge	0.702	508
16	Silverpine Forest	0.541	258
17	Stranglethorn Vale	0.722	188
18	Swamp of Sorrows	0.488	154
19	The Hinterlands	0.711	793
20	Tirisfal Glades	0.699	406
21	Western Plaguelands	0.517	230
22	Westfall	0.735	330
23	Wetlands	0.584	196

of the region, among others. Unfortunately we could not find any relationship.

The Figure A.8 shows the transition probability matrix between the 23 states of the process. The columns with major number of positive probabilities are 9, 14 and 19, which are the larger merchant zones in Eastern Kingdom. In other words, there are a high number of displacements for these zones. Suppose a character is in Elwynn Forest, a merchant zone with level range from one to ten. The most likely zones for which he will advance are Dun Morog (level 1-10), Stranglethorn Vale (level 25-35) and Duskwood (level 20-25). Suppose he advances to Dun Morog, also a merchant zone with same level range. The most likely zones for advance are Elwynn Forest, Loch Modan (level 10-20) and Wetlands (level 20-25). If he moves for Loch Modan, there is a high probability that his next advance will be for Badlands (level 45-48) or Dun Morog. Observing Figure 2a, we can see that Dun Morog, Loch Modan, Wetlands and Badlands are geographically proximate. Repeating this analysis in another zones, we find that the majority of character movement in Eastern Kingdom map occurs between zones geographically proximate and zones with proximate level range. Unfortunately, we did not have information about which zones are related through quests, and we can not analysis the probabilities between these zones.

Table 5 presents an estimate for the invariant distribution of the process. The invariant distribution can be interpreted as the probability of the process be in each state, after its reach the equilibrium. So then, zones with higher probability generate more load in the game server. The highest probability, 0.202, was observed in Elwynn Forest. This is the zone with the major merchant center of Eastern Kingdoms, and also is the most popular zone between players. The other three major merchant zone in Eastern Kingdom, Stranglethorn Vale, Trisfall Glades and Dun Morog, showed the second, third and fifth highest probabilities, respectively. The lowest one, 0.006, was observed in Searing Gorge, which is the locked zone for Aliance characters. According these results, the load generate in the game server by Elwynn Forest is much higher than in other zones.

5.1. Simulation Results

In the following analysis, we simulated 100 data sets with 178 simultaneous players, the average number of simultaneous players in the sample, and 24 hours of game for all of them. The density of the time spent in the states was estimated by the Weibull distribution with the MLE estimators and by Kernel with the bandwidth given by the methods of Sheather and Jones (1991), Silverman (1986) and unbiased cross-validation.

Table 6 shows the Kolmogorov-Smirnov test results for the times simulated by kernel with Silverman, Cross-Validation and Sheather & Jones bandwidth and by the

Table 5: Estimate for the invariant distribution of the process.

State	Zone	Probability
1	Alterac Mountains	0.012
2	The Hinterlands	0.009
3	Arathi Highlands	0.006
4	Badlands	0.005
5	Blasted Lands	0.031
6	Burning Steppes	0.032
7	Wetlands	0.015
8	Deadwind Pass	0.007
9	Dun Morogh	0.043
10	Westfall	0.039
11	Duskwood1	0.046
12	Western Plaguelands	0.006
13	Eastern Plaguelands	0.008
14	Elwynn Forest	0.378
15	Tirisfal Glades	0.063
16	Hillsbrad Foothills	0.027
17	Swamp of Sorrows	0.024
18	Loch Modan	0.008
19	Stranglethorn Vale	0.166
20	Redridge Mountains	0.025
21	Searing Gorge	0.004
22	Silverpine Forest	0.010
23	Duskwood2	0.037

Weibull distribution. The presented values are the mean of the 100 values obtained from each simulated data set. Here, the null hypotheses of the test says that the sample and simulated times are drawn from the same distribution. In parentheses are the p-values of the tests. As we can see, the only method showing p-value lower than 0.05 in all states was the Sheather & Jones method. Therefore the Sheather & Jones method returns bandwidth values more suitable to our data. Comparing the parametric and nonparametric approaches, we found that the parametric shows lower statistics in 14 of the 23 states.

Table 6: Kolmogorv-Smirnov test for the simulated times.

State	Zone	Silverman	Cross-Validation	Sheather & Jones	Weibull
1	Alterac Mountains	0.228(0.052)	0.133(0.567)	0.164(0.312)	0.127(0.618)
2	The Hinterlands	0.308(**0.007**)	0.264(**0.033**)	0.216(0.126)	0.156(0.463)
3	Arathi Highlands	0.238(0.098)	0.155(0.538)	0.213(0.177)	0.124(0.785)
4	Badlands	0.315(**0.022**)	0.153(0.651)	0.243(0.145)	0.135(0.796)
5	Blasted Lands	0.109(0.088)	0.044(0.948)	0.050(0.884)	0.072(0.503)
6	Burning Steppes	0.269(**0.003**)	0.126(0.467)	0.166(0.169)	0.140(0.335)
7	Wetlands	0.286(**0.008**)	0.143(0.494)	0.204(0.127)	0.132(0.596)
8	Deadwind Pass	0.127(0.400)	0.071(0.953)	0.116(0.517)	0.137(0.329)
9	Dun Morogh	0.119(**0.024**)	0.069(0.462)	0.063(0.561)	0.091(0.161)
10	Westfall	0.175(**0.019**)	0.058(0.932)	0.069(0.837)	0.102(0.411)
11	Duskwood1	0.135(**0.003**)	0.100(0.064)	0.085(0.166)	0.077(0.254)
12	Western Plaguelands	0.184(0.324)	0.111(0.878)	0.161(0.513)	0.130(0.774)
13	Eastern Plaguelands	0.150(0.617)	0.100(0.928)	0.094(0.946)	0.152(0.603)
14	Elwynn Forest	0.071(**0.002**)	0.029(0.628)	0.028(0.669)	0.055(0.053)
15	Tirisfal Glades	0.068(0.614)	0.060(0.753)	0.063(0.693)	0.079(0.441)
16	Hillsbrad Foothills	0.205(**0.003**)	0.159(**0.049**)	0.096(0.496)	0.064(0.899)
17	Swamp of Sorrows	0.095(0.141)	0.039(0.970)	0.045(0.912)	0.083(0.274)
18	Loch Modan	0.269(**0.023**)	0.261(**0.031**)	0.240(0.061)	0.179(0.290)
19	Stranglethorn Vale	0.132(**0.001**)	0.037(0.871)	0.036(0.876)	0.075(0.167)
20	blackridge Mountains	0.151(0.078)	0.095(0.539)	0.119(0.269)	0.099(0.492)
21	Searing Gorge	0.490(**0.004**)	0.221(0.536)	0.287(0.250)	0.215(0.578)
22	Silverpine Forest	0.201(0.112)	0.088(0.918)	0.144(0.447)	0.109(0.759)
23	Duskwood2	0.178(**0.000**)	0.099(0.063)	0.094(0.088)	0.068(0.397)

Table 7 shows the times spent to simulate the movement dataset for 75, 150, 300 and 600 simultaneous players in different game times using Weibull distribution to simulate the times spent in the states of the process. We used the parametric approach considering not everyone have a sample available.

Table 7: Time spent in the simulations.

		Session time (h)		
		12	24	36
Number of players	75	19s	21s	26s
	150	22s	31s	39s
	300	28s	55s	84s
	600	53s	123s	221s

6. Conclusion

We proposed to model the character movement in the game map as a jump process. We found that for our data the time spent in the map zones was better modeled by the Weibull distribution with parameters estimated by the MLE. Furthermore, we showed that is possible to model the time spent in the zones by Kernel Density Estimation.Through the proposed methodology, we could generate new samples of the mobility game, with different numbers of players and session times, showing behavior similar to the observed data. Therefore, we conclude that the jump process can represent reliably the movement of characters in the game. Furthermore, the data showed that some zones have charge traffic much higher than others. That information can be used to improve the game servers architecture. As future work, would be interesting defining the states of the process as smaller regions inside the zones, as cities, farms, campings, castles, among others regions defined in the game. This would allow to accomplish a more detailed analysis of the mobility.

References

V. Lehdonvirta, M. Ernkvist, 2011. Converting the Virtual Economy into Development Potential: Knowledge Map of Virtual Economy, Washington, DC; infoDev/World Bank. Available at http://www.infodev.org/publications, 2011.

K.-T. Chen, C.-Y. Huang, P. Huang, C.-L. Lei, An Empirical Evaluation of TCP Performance in Online Games, in: Proceedings of ACM SIGCHI ACE 06, Los Angeles USA, 2006.

P. Svoboda, W. Karner, M. Rupp, Traffic Analysis and Modeling for World of Warcraft, in: Communications, 2007. ICC2007. IEEE International Conference on, 1612 –1617, 2007.

Y. Fang, L.; Guotao, Z. Wenli, Traffic recognition and characterization analysis of mmorpg, in: International Conference on Communication Technology (ICCT), 2006.

C.-A. La, P. Michiardi, Characterizing User Mobility in Second Life, in: WOSN, 2008.

M. C. Santos, análise Espaço-Temporal da Mobilidade em Jogos On-Line Massivos, 2010.

S. A. Tan, W. Lau, A. Loh, Networked game mobility model for first-person-shooter games, in: Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games, NetGames2005, ACM, New York, NY, USA, 1–9, 2005.

J. L. Miller, J. Crowcroft, Avatar movement in World of Warcraft battlegrounds, in: Proceedings of the 8th Annual Workshop on Network and Systems Support for Games, NetGames2009, IEEE Press, Piscataway, NJ, USA, ISBN 978-1-4244-5605-5, 1:1–1:6, 2009.

P. Ferrari, A. Galves, construction of Stochastic Processes, Coupling and Regeneration, XIII Escuela Venezolana de Matemática, 2000.

E. H. Feng, continuous Time Markov Processes, 2004.

J. A. Hartigan, M. A. Wong, A K-Means Clustering Algorithm, Applied Statistics 28 (1979) 100–108.

P. Hall, Large sample optimality of least-squares cross- validation in density estimation, Ann. Statist (11) (1983) 1156–1174.

S. J. Sheather, M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society, Series B, Methodological 53 (1991) 683–690.

B. Silverman, Density estimation for statistics and data analysis, Monographs on Statistics and Applied Probability, Chapman and Hall, ISBN 9780412246203, 1986.

D. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley series in probability and mathematical statistics: Applied probability and statistics, Wiley, ISBN 9780471547709, 1992.

S. J. Sheather, Density estimation, Statistical Science (2004) 588–597.

W. Conover, Practical Nonparametric Statistics, John Wiley & Sons, ISBN 9780471168515, 1971.

Appendix A. Figures

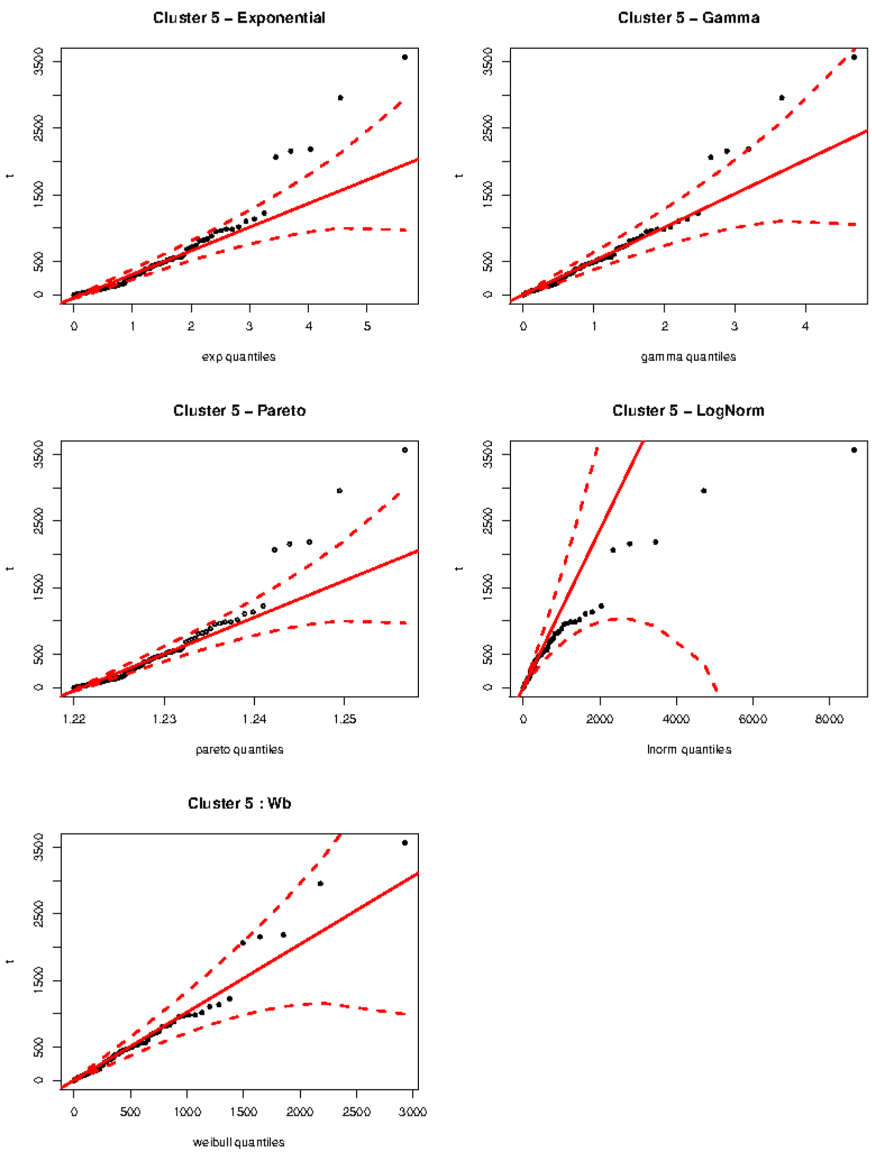


Figure A.5: QQ-plot of the times spent in the state 5.

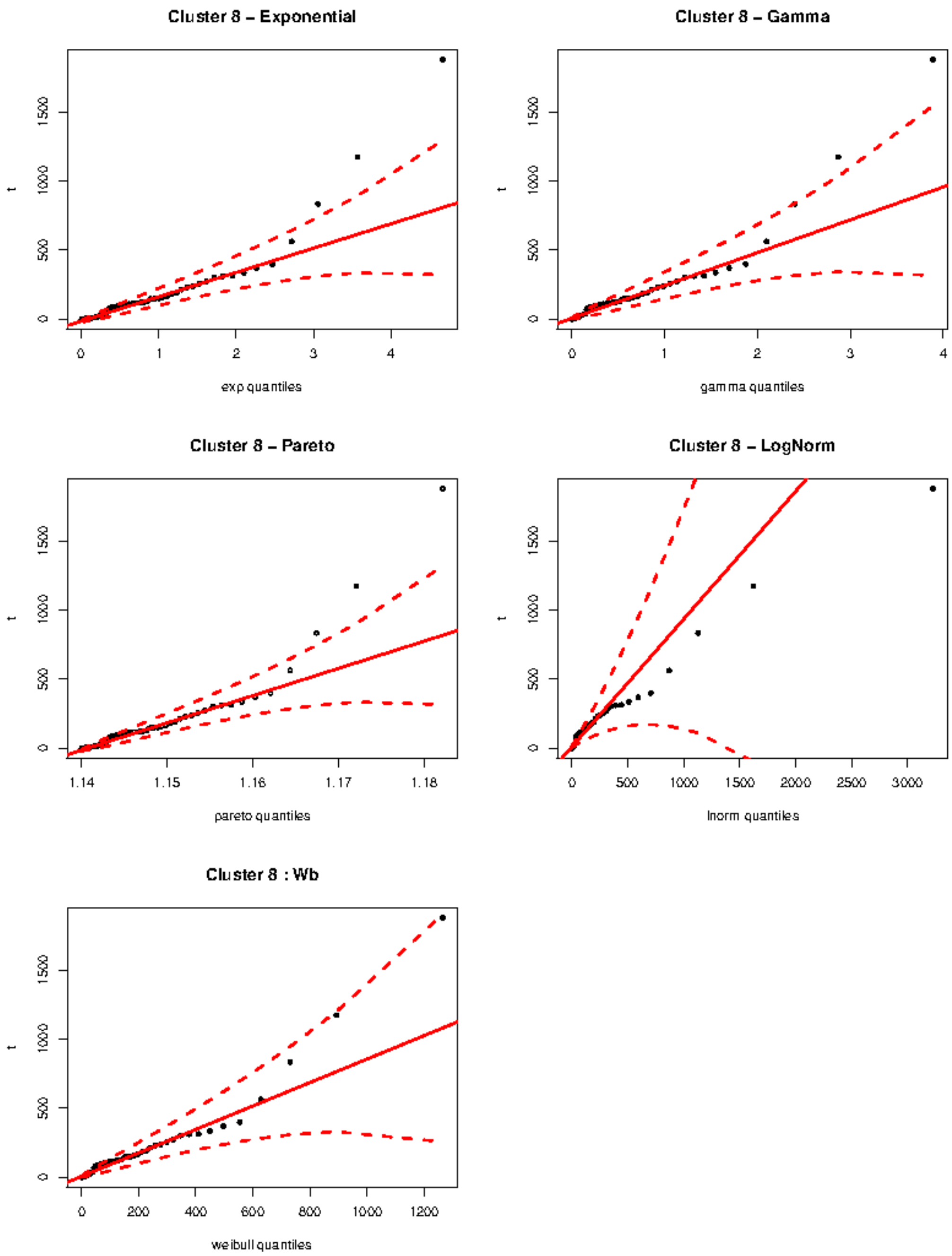


Figure A.6: QQ-plot of the times spent in the state 8.

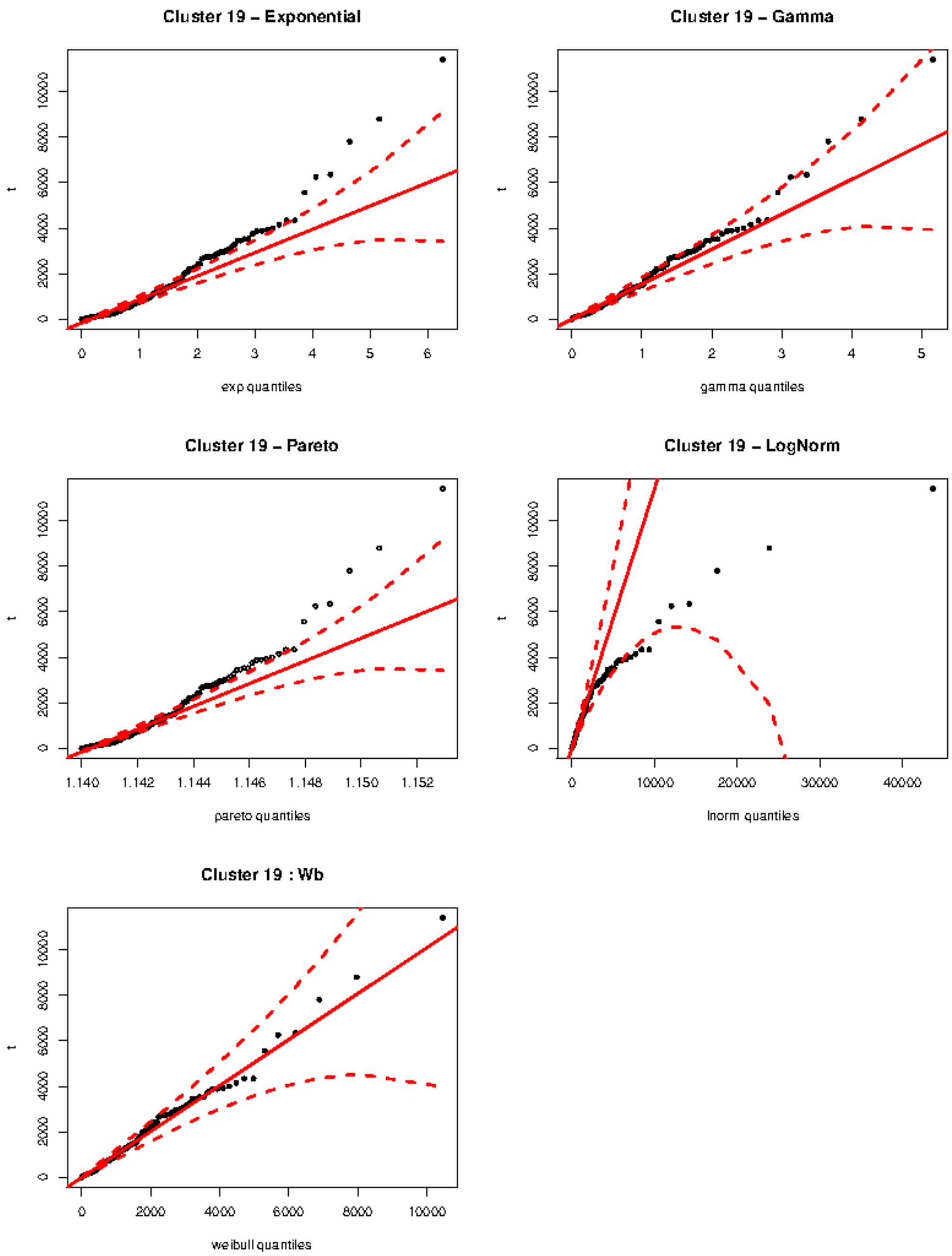


Figure A.7: QQ-plot of the times spent in the state 19.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	0	0	0	0	0	0	0	0	0	0	0.054	0	0	0.135	0.757	0	0	0	0	0	0.054	0
2	0	0	0.161	0	0	0	0.065	0	0	0	0	0.065	0.065	0.194	0.097	0.194	0	0	0.097	0	0.065	0	0
3	0	0.192	0	0	0	0	0.269	0	0.077	0	0	0	0	0.077	0.077	0.269	0	0.038	0	0	0	0	0
4	0	0	0	0	0	0.118	0	0	0.353	0	0	0	0	0.059	0	0	0	0.412	0	0	0.059	0	0
5	0	0	0	0	0	0	0	0.011	0	0	0.011	0	0	0.138	0	0	0.798	0	0.032	0	0	0.011	0
6	0	0	0	0.083	0.028	0	0	0	0	0	0	0	0	0.361	0	0	0.056	0	0.083	0.25	0.139	0	0
7	0	0	0.32	0	0	0.04	0	0	0.16	0	0	0	0	0.16	0	0.04	0	0.28	0	0	0	0	0
8	0	0	0	0	0.04	0	0	0	0.02	0	0.18	0	0	0.12	0	0	0.56	0	0.04	0.02	0	0	0.02
9	0	0	0.018	0.045	0	0.009	0.108	0	0	0.009	0.009	0.009	0.054	0.577	0	0.027	0	0.117	0	0.009	0.009	0	0
10	0	0	0	0	0.016	0	0	0.016	0	0	0	0	0	0.548	0	0	0	0	0.097	0	0	0	0.323
11	0	0	0	0	0.012	0	0	0.22	0	0.006	0	0	0	0.243	0	0.012	0.023	0	0.179	0.012	0	0	0.295
12	0	0.083	0	0	0	0	0	0	0	0	0	0	0.292	0.125	0.333	0.125	0	0	0.042	0	0	0	0
13	0	0.364	0	0	0	0	0	0	0	0	0	0.455	0	0	0.091	0	0	0	0.091	0	0	0	0
14	0	0	0	0	0.002	0.034	0.01	0.005	0.185	0.11	0.139	0.002	0	0	0.002	0.014	0.065	0.005	0.163	0.106	0	0.002	0.156
15	0.09	0.045	0	0	0.015	0	0	0	0	0	0	0.209	0.015	0.015	0	0.149	0.015	0	0.104	0	0	0.343	0
16	0.437	0.141	0.127	0	0	0	0	0	0.056	0	0	0.014	0	0.056	0.113	0	0	0	0	0	0	0.056	0
17	0	0	0	0	0.766	0.014	0	0.041	0	0	0.034	0	0	0.09	0	0	0	0	0.028	0.028	0	0	0
18	0	0	0	0.4	0	0	0.167	0	0.367	0	0	0	0	0.033	0	0	0	0	0	0	0.033	0	0
19	0	0.006	0	0	0.012	0.042	0.006	0.006	0.018	0.03	0.204	0	0.006	0.377	0.048	0.006	0.024	0	0	0	0	0	0.216
20	0	0	0	0	0	0.2	0	0.015	0.015	0	0.077	0	0	0.585	0	0	0.077	0	0.031	0	0	0	0
21	0	0	0	0	0	0.364	0	0	0.182	0	0	0	0	0.273	0	0.091	0	0	0.091	0	0	0	0
22	0	0	0	0	0.061	0	0	0	0	0	0	0	0	0.03	0.606	0.273	0	0	0	0.03	0	0	0
23	0	0	0	0	0	0	0	0	0	0.07	0.343	0	0	0.436	0	0	0	0	0.151	0	0	0	0

Figure A.8: Transition probability matrix between the 23 states of the process.