UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

APPROXIMATION OF THE DISTRIBUTION OF THE SUM OF CORRELATED RANDOM VARIABLES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

PABLO IGNACIO SARMIENTO ROJAS

PROFESOR GUÍA:
FERNANDO ORDOÑEZ PIZARRO

MIEMBROS DE LA COMISIÓN:
DENIS SAURÉ VALENZUELA
FELIPE LAGOS GONZALEZ

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: PABLO IGNACIO SARMIENTO ROJAS
FECHA: 09/12/2019
PROFESOR GUÍA: FERNANDO ORDOÑEZ PIZARRO

APPROXIMATION OF THE DISTRIBUTION OF THE SUM OF CORRELATED RANDOM VARIABLES

Al despachar vehículos de emergencia, es importante considerar no solo los tiempos esperados de los caminos por los que viajan, sino que también es relevante considerar la estructura que tienen estos caminos, ya que no es lo mismo calcular el valor esperado de una ruta con un intervalo de confianza mayor o menor. Para conocer la estructura de los caminos es interesante estudiar cómo obtener la distribución de los tiempos de viaje que constituye un camino, la idea es obtener esta distribución a través de la suma de las distribuciones de los arcos que lo componen, representándolo en un grafo dirigido, con sus respectivas correlaciones, dado que es un comportamiento mas realista que debe considerarse, como se observa en el trabajo de [13], las correlaciones mejoran el rendimiento para modelar el comportamiento de una ruta.

Anteriormente, se estudió en [6] cómo aproximar la suma de variables aleatorias correlacionadas, mediante el uso de la función generadora de momentos en distribuciones lognormales, obteniendo buenos resultados, sin embargo, el desarrollo de esta herramienta está contenido en el mundo de distribuciones lognormal, es por ello que una nueva aplicación de esta idea se desarrolla aqui para aproximar la suma de variables aleatorias correlacionadas, en este contexto se estudia esta aplicación en distribuciones gamma. Esta nueva aplicación en distribuciones gamma se desarrolla utilizando datos simulados, donde se obtienen excelentes resultados,

Este enfoque se desarrolla posteriormente mediante el uso de una estructura de datos reales. Como es posible que los datos reales no sigan una distribución gamma propiamente tal, se eligieron aquellos datos que mantienen la estructura deseada que se está desarrollando en el trabajo, suponiendo básicamente que se tiene un conjunto de datos reales que se distribuyen como queremos, es decir, una distribución gamma multivariada, esto permite verificar que teniendo una estructura real de datos que siguen una distribución gamma, entonces la forma de reconstruir la distribución de tiempos de viaje elaborada aquí es adecuada. Nuevamente se obtienen prometedores resultados.

Finalmente, lo anterior conlleva a tener una herramienta disponible que, bajo ciertas condiciones muy generales, permite calcular la distribución de la suma de variables aleatorias correlacionadas de una manera funcionalmente muy efectiva.

Como trabajo futuro, se recomienda comenzar a agregar aplicaciones de este trabajo con métricas de aversión al riesgo, lo que está en línea con las motivaciones iniciales del desarrollo de esto, lo que nos permitiría tener una visión más completa de la distribución de tiempo de viaje y avanzar su desarrollo a algo aplicable en la vida diaria.

ABSTRACT OF THE MEMORY TO OPT
THE TITLE OF INGENIERO CIVIL INDUSTRIAL
BY: PABLO IGNACIO SARMIENTO ROJAS
DATE: 09/12/2019
SUPERVISOR: FERNANDO ORDOÑEZ PIZARRO

APPROXIMATION OF THE DISTRIBUTION OF THE SUM OF CORRELATED RANDOM VARIABLES

When dispatching emergency vehicles, it is important to consider not only the expected times of the roads they travel, but it is also relevant to consider the structure that the roads have, since it is not the same to calculate the expected value of a path with a greater or shorter confidence interval. To know the structure of the paths it is interesting to study how to obtain the distribution of travel times that constitutes a path, this distribution is interesting to obtain through the sum of the edges that make up this path, when representing this structure in a graph, with their respective correlations since it is a natural behavior that must be considered to be more realistic, as we seen in the work of [13], correlations improve performance in order to model the behavior of a path.

Previously, it was studied in [6] how to approximate the sum of correlated random variables through the use of the Moments Generating Function which was tested in log-normal distributions with good results, however the development of this interesting tool is contained in this world of log-normal distribution, that is why a new application of this idea is developed in this way to approximate the sum of correlated random variables, as long as it can be applicable in other contexts and applying it in gamma distributions, worked entirely in a first instance through simulations, for explanatory purposes of the generalization raised, obtaining excellent results.

This approach is subsequently developed with gamma distributions through the use of a real data structure. As it is possible that the real data does not follow a gamma distribution, those that maintain the desired structure that is being developed at work were chosen, assuming basically having a set of real data that is distributed as we want, that is, a multivariate gamma distribution, this allows us to check if before a real structure of data that will not be exactly the distribution we want, rather roughly, this way of reconstructing the distribution is adequate. Here you also get excellent results.

Finally, a tool is available that, under certain very general conditions, makes it possible to calculate the distribution of the sum of correlated random variables in a functionally very effective way.

As a future work, it is recommended to start adding applications of this work with risk aversion metrics, which is in line with the initial motivations of the development of this tool, which would allow us to have a more complete vision of the distribution of the travel time.

Acknowledgements

I want to thank my family for the unconditional support they gave me throughout my career and also want to thank my girlfriend Melissa for being with me and supporting me in all the processes that I have had to live.

It is also very important to thank my professor, Fernando, for the help, support and patience he gave me throughout my work.

Finally, I would like to thank the entire team of Firefighters from the 518 office that was an important part of the development of all the learning that allowed me to reach this moment.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivating Problem

When dispatching emergency vehicles it can be critical to consider the travel time distribution of different possible trajectories, because each second of delay in the arrival potentially may cost lives or cause irreparable damage to both people and material goods that surround them. it is important at the time of dispatch to consider not only average travel times, but also the risk associated with selecting each path. In this context this refers to be able to make use of the distribution of the random variable representing the travel time over a path of the random variable representing the travel time over a path, when selecting its risk aversion.

Having the ability to consider travel time risk at the moment of dispatch creates another problem, that is how to store the large amount of data that involves all the paths in a city. If we think about it at street level, each arc has an underlying distribution in its travel times, (for example each arc would have to store the type of its distribution it and the distribution parameters). However, trying to store travel time distributions for the paths composed from these edges constitutes a quantity of information at a certain scale that is difficult and expensive to manage. The observation that storing the path travel times is impractical due to size raises the question of whether the, much smaller, arc travel time distribution data can be used efficiently to compose the path travel time distributions. very important when considering that the problem becomes more complex when thinking about distributions with correlations, which makes interacting between them not so direct as opposed to considering them independent,

To address this problem we proceed as follows

1. Consider a directed graph and get the distribution of the travel time over every edges
2. Get the travel time correlations between edges
3. Develop a method to compute the complete travel time distribution over a path from the distribution of the underlying edges and their correlations

This approach provides the travel time distribution over arbitrary paths in a graph without having to store that information for every path.

1.2 Problem Statement

To develop an algorithm for risk averse shortest path, its important to consider the distribution of the possible paths, because that not only gives the information about the expected travel time but also information about structure of the travel time, that is, the probability that the travel time be in a certain interval and from here develop algorithms that consider this elements for a more robust travel time estimation that use elements as risk measure. So for the situation described above, it is a good idea to consider the information of distribution of all paths , which is possible for small graph. However, a reasonable graph representing the street network of Santiago has more than 650,000 directed edges so it is impossible to store in this case the information for all possible paths, but we can store the travel time distribution for the graph edges.With this information, the travel time distribution over a given path is obtained by the convolution of the distributions over the edges of the path. To make this possible generalize a the method developed in [6]. This method makes the following assumptions (points 1 and 2) and needs (points 3 and 4):

1. The travel time of every edge, followings the same distribution
2. There is a known distribution that approximates the distribution of the sum of edges distribution. For example the sum of log-normal can be well approximated by a log-normal distribution
3. The distribution of the sum of random variables, considering their correlations
4. A quadrature to approximate the integral of the distribution the integral of the distribution to which the sum of the travel times of the edges approaches.

So, from the work developed in [6] we have an important structure to work but we want to extend the method because that work is limited to the sum of log-normal distributions. We aim to extend this methodology for the use of other distributions, of if we do not know the distribution to approximate the path travel time.

Also the work in [6] develops an important methodology, the Equivalent sum of Log-normal method (ESL), which computes a large amount of convolutions by a montecarlo simulations. So it is a powerful methodology in which we add value extending it for use under different assumptions.

From all the above, the problems that it sought to solve are:

1. Reconstruct the distribution of the travel time in paths using the information of the edges and the correlation between them. This is because is not possible store all distributions of all paths in a graph
2. Found other distribution of travel time in the edges to apply this methodology and develop a way to find a good distribution every time you need
3. Recognize how the previous work changes in its procedures when using new distributions and what implications does that have at the logic level that was developed to approximate the sum of travel times.

1.3 Objectives

1.3.1 General Objectives

The objectives outlined below originate to address the question: Is it possible to approximate the distribution of travel times of a path, from the distribution of the travel times of the edges that compose it and the correlations between them in a general procedure but applied to a specific distribution (in this case, the gamma distribution)?

1. Elaborate a methodology to replicate the estimation of the distribution of travel times of a path, from the distribution of the travel times of the edges that compose it and the correlations between them with another distribution, different from the Log-Normal
2. Reconstruct the distribution of travel times of a path, by using the distribution of travel times of the edges that make it up and the correlation between them, in a graph built with real data from Transantiago

1.3.2 Specific Objectives

This work is organized along the following specific objectives:

1. To get requirements that are needed to be able to apply the estimated travel times
2. Develop an analogous methodology to "Equivalent sum of Log-normal method" for other distributions, that is, apply the essence of the Monte Carlo simulation in ESL but for another distribution, other than log-normal,
3. Validate proposed methodologies with Log-Normal distributions (and compare them with those made in previous studies)
4. Sample probability distributions from a density or distribution function, different of a Log-Normal to test that analogous methodology
5. Measure approximation adjustment of new methodology with Kolmogrov Smirnov test for different parameters of the distributions
6. Measure goodness of fit of this methodology using a real travel times from Transantiago information and a real structure of edges

1.4 Process to Follow

The structure underlying the engineering thesis is constituted as follows:

1. Analyze what are the necessary inputs of the method [6] and subsequently establish the new input requirements for the new application.
2. Generate a proposal to find other applications of the method based on the needs raised and how to solved its
3. Once this is established, apply with a Log-Normal distribution to be able to compare the effectiveness of the proposed method with the methodology developed in [6]
4. Carry out the propose another distribution, that meets necessary assumptions
5. Validated sample, apply the method in Transantiago data, this is:
   (a) Project the Transantiago data in a directed graph
   (b) Select the paths on the graph to analyze
   (c) Identify the travel time distribution to the data on selected paths
   (d) Apply the method N times in a random set of paths
   (e) Calculate adjustment metrics to the roads according to the size

This work constitutes an extension to the method introduced in [6], conducts tests with Log-Normal distributions, similar to the work in [6], then tests it with other distributions and finally apply it and test it with real data.

The steps above are presented in this document in the following sections:

First in section 3.1 and 3.2 points 1, 2 and 3 are presented. This Chapter has the objective to describe the method developed in [6] and how to apply in other context. This change in the original method is tested using the Log-Normal distribution to make sure that does not change the results found previously. In section 3.3 and 3.4 points 4 and 5 of the process are introduced. Here the focus is to apply the proposed new method with different distribution, and set all the things required for this application. In section 3.4 presents point 6, which focuses in use real data to apply the modifications of the approximation of travel time of paths. The main results are in the Chapter 3. Finally the conclusions and next steps to this research line are in Chapter 4.

2 Background Theory

2.1 Definitions

Convolution: Mathematical function between two functions (f * g) that "blends" these functions to produce a third function that expresses the amount of overlap of one function g as it shifted over another function f, The domain it depends of the domain of the functions that are participating. A convolution is defined by the formula as

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)\mathrm{d}\tau$$

Simulation: In this context, simulation it will be used to describe the generation of set of data that follows a specific probability distribution.

Moment Generating Function (MGF): Is a function that express an alternative specification of a probability distribution. The MGF of a random variable X is defined by the formula below

$$M_X(t) = E[\mathrm{e}^{tX}] = \int_{-\infty}^{+\infty} \mathrm{e}^{tX} f_X(x)\mathrm{d}x$$

Kolmogrov-Smirnov Test: It is a non-parametric test that measures the goodness of fit between two continuous unidimensional probability distributions and unidimensionals.

Sample: In this context, given a probability distribution, a sample of size N is a set of N random variables identically distributed according to that probability distribution.

non linear equation: Any equations that cannot be written as

$$f(x) = 0$$

, where the function f is not affine

Affine function: is a linear function plus a translation

system of nonlinear equations: Is a set of equations in which at least one equation is non linear.

Mean square error: It a measure of goodness of fit that consist of the average of the difference squared between the estimator and what you want to estimate.

Quadrature: In this context, is a numerical approximation of an integral A **variable**, in this context, is a value that can takes different values in different situations, so when is a **random variable** that means a variable that comes from a random phenomenon and the values that takes can be continuous or discrete. In this context, this phenomenon follows a **probability distribution**. This concepts describes the probabilities of the possible values that the random variable can take. If the random variable is discrete, then the probability distribution describe the probability of each value than the random variable can take, if is continuous then the probability is describe as the area under the curve of its **probability density distribution (PDF)**. A PDF is a function that at any given point gets the relative probability that a random variable take that value. In continuous random variable the PDF in a specific value could be different of zero but the probability that the variable takes that exact value is zero since there are infinite set of different points so the volume of the interval is zero, but is commonly used to describe the probability that a random variable fall into a particular range of values.

The **likelihood** in essence, for a set of data corresponds to the value in probability of a distribution, for a set of given parameters, of obtaining that set of data, so from this point the **Maximum Likelihood Estimation** or **MLE** corresponds to the set of parameters for a distribution that are most likely to have obtained a given set of data. This is achieved by assuming that the data set are the most likely to have been obtained.

The joint distribution will be sampled using the method of simulation Acceptance- Rejection. This alternative sampling method is useful if it is not possible to use the traditional method of the inverse of a distribution of random variable. This method can be used to simulate a continuous or discrete random variable. Here case we only explain the continuous case but the discrete case is very similar. This method was described in detail in [14]. Similar to that reference, we start assuming that we know the density f(x) of the distribution F, the distribution that we want to sample. The basic idea at this point is to find an alternative distribution G, with density g(x) that meets two conditions:

1. The distribution G has an efficient method to be sampled
2. The density g(x) is a good approximation to the density f(x). In particular it is described in [14] that g (x) is a function which is bounded by a constant c>0; $sup_x f(x)/g(x) \leqslant c$. It is desirable that c as close to 1 as possible.

Given this notation, the method Acceptance-Rejection to sample distribution F is described below::

1. Generate a random variable Y distributed as G
2. Generate U (independent of Y), as an Uniform[0,1]
3. if

$$U \leqslant \frac{f(Y)}{cg(Y)}$$

; then set *X*=*Y(accept)*; otherwise go back to 1 (reject)

This method is described in [14], for more examples and a detailed description.

Maps are usually represented using graphs. A graph, in a simple way, is a set of nodes joined by links called edges. There are two types of graphs, first directed graphs, in which the edges between nodes indicate the direction of the relationship, that is, if node A is related to node B or node B to node A or both. Second with undirected graphs, in them the edges only indicate which node is related to which, without detailing the direction of the relationship,which can be in any direction. In this work, directed graphs will be used.

2.2 Travel Time Distribution

The analysis of travel time distributions has been widely studied. It is specially important in the context of emergency vehicles, where there are more variables to consider than the expected travel time, such as variability or to risk aversion constraints (for example, to arrive with no more than 2 minutes of delay). That is why it is an important line of research, in which it is usual to consider only the expected travel time as an estimate of the travel time, but in this way, the structure of this expected travel time, that is, to consider the travel time as a random variable and the set of possible times as a probability distribution in this random variable, hence the importance of considering those elements.

The work in [12] studies about the reliability of travel time of private car journey in Melbourne Australia, conclude the use of travel time distribution is a better approach to study than considering only the deterministic. In, [13] the authors work in algorithms to improve the reliability of travel time, considering normal distribution and correlated link travel times, which is a good approach seeing the results the improve in this study is to add real-time data to improve the current distribution of the paths. This is what be studied in [4], this study uses real time data to the travel time considering the change in the structure of their distribution in time.

In summary, its important consider the dynamic structure of the distribution of travel time and the correlations between them to better estimate travel times over the graph, in particular their deviation and risk.This work focuses on a method to compute the travel time distributions over arbitrary paths in the network, considering the distributions of the edges that conform a complete path and the correlations between them.

2.3 Sum of dependent random variables

Although in this context we study the sum of correlated or dependent random variables to reconstruct the de-distribution of said sum to represent a route in a graph, there are studies in other contexts that have been responsible for studying this same topic. Then we see the work done by [10], where he explains some results about the probability that the sum of n sub-exponential dependent variables exceeds a threshold μ. It also performs an interesting analysis on the effect of the tails of one of the distributions that add up. Here [10] performs interesting analyzes to explore the sum of variables, however the approach moves away from the analysis performed here, rather it is focused on analysis, for example, of risk.

It is also possible to consider the work done by [7] which the distribution of the sum of the dependent random variables is calculated by means of Computing the distribution of the sum of dependent random variables via overlapping hypercubes, however this methodology focuses mainly on the calculation of the VaR, testing the convergence of the geometric algorithm, which is alternative to Monte Carlo estimation methods. Here, like the previous one, the work focuses on methods oriented towards finance, which moves away from the work developed here.

Finally, to mention another work in this area of the sum of random variables, we can mention the work of [3], here it is more similar to what is presented in this work, since it is sought to calculate the distribution of n dependent random variables given its joint distribution, however, it is a way that moves away in its methodology to what is presented here, looking more graphically for this adjustment, which is a different way although it's based on inputs similar to those used in this work, so then in the future it could be used to contrast results if possible.

Given the above, in general, studies are focused on financial issues, given its high implication in the distribution of returns and also in some other areas, but focusing efforts on different methodologies than those presented here.

2.4 Previous work

2.4.1 Computing Path Travel Time Distributions

The present thesis is an extension of the work develop by [6] so its important to describe their work to get the complete framework.

In the work of *Felipe Lagos*[6], work with a graph in which the travel times are represented by a random variable in every edge following a log-normal distribution, they also consider the correlation between them. This information is used to compute the travel time distribution of arbitrary paths in a more realistic situation (stochastic travel times and correlation, that is a natural structure for a graph that represents travel times instead of deterministic structure, considering only expected travel time). It is important to mention that this previous work assumes that the same distribution, albeit with different parameters, is considered on every edge of the entire path. This is a strong assumption that we will also consider in the present work. It's important remark that, from a technical point of view, it is important to maintain this assumption because adding correlated random variables that do not follow the same distribution moves away from the previous work on which we want to contribute, but from a practical point of view, in real life the adjacent roads or belonging to the same road component, it is reasonable to consider that they could have the same structure, that is, that their distributions would behave similar, in a certain way, so While it is a strong assumption, it is not a crazy thing.

The issue is try to solve is the estimation of the distribution of travel time on a path , considering the structure mentioned above about distribution in edges and correlation between them.In short, to solve this they use two main things:

1. First, it is possible to write the travel time distribution in two different ways:
   - As the sum of every random variable that makes up the the path, that is, the convolution of the random variables and their correlations
   - As a single distribution that represents the travel time over the entire path
2. A property that states: if two distributions have the same moment generating function, then they are identical at almost all points [8]

With these things, its possible to construct an equation system equating the MGF of the sum of the random variable of the path and the MGF of the travel time distribution along the path. This equation system is presented in the next equation for the case of use a Log-normal, have two parameters, giving:

$$\Psi_y\left(S_{\mathrm{i}}, \mu_x, \sigma_x\right) = \Psi_{\left(\sum_{k=1}^{k} Y_k\right)}\left(S_{\mathrm{i}}, \mu, C\right) \text{ for i=1,2 [1]} \tag{2.1}$$

Remembering that if we solve this system, then the two distribution underlying of the MGF are equal in almost all points, so we could use as distribution of the path the sum of the edges or the single distribution that we just approximate.

The left side of the equation (2.1), represents the distribution of a single random variable that represents the travel time of the path, that is logical following the idea that any path is possible to represent like that independent if this path is formed by edges. The right side of (2.1) represents the sum of the random variables Y_k, here is using the idea that the complete path is formed by edges and naturally the sum of the distribution of this edges form the distribution of the complete path, you have to keep in mind that these edges have significant correlations to consider in this sum. The two values of i comes from the different values that the parameter S of the MGF can ake to generate the number of equations that are needed to determine the unknowns in this system of equations, namely the expected travel time and standard deviation of the path distribution. That is, for each variable that we need to get, in this case only the mean μ and the standard deviation σ of the left side, we have a single value of the parameter S, in this case 2. It is true that one might think that only one parameter needs to be calculated since, having the distributions of the edges underlying the path, one could calculate the mean of the distribution of the path by adding the means of the edges, however, the work itself It consists in finding the best approximation, based on MGF, of the proposed distribution, which being real data doesn't necessarily imply that it distributes as such, but rather certain dynamics were found that allow us to model the problem as described and It will be described later, therefore, probably approximating the mean by adding the means of the edges would obtain a result that is at least acceptable, it would not be the best possible parameter to find since in a way the actual data does not distribute as a log-normal proper. Rather, you should find a distribution that is as similar to the distribution of the path, which leads us to move forward with this methodology.

To get this value "S" we consider that the MGF for a variable X is an expected value function of $\exp(-sX)$, this is in essence a weighting factor for the pdf, hence, in order to find good parameters we need to put attention to this function. If we choose a very big value for s, we will probably overweight the tail of the distribution. On the contrary, small values might only take into account the head portion. To avoid these situations, we develop a way to find good s values. The mass center of the distribution L is:

$$\frac{1}{2s} = \int_0^L \exp(-sx)\mathrm{d}x = \frac{1 - exp(-sL)}{s} \tag{2.2}$$

Here we use $\frac{1}{2s} = \frac{I}{2}$, because it is half of the total area for the integral for $\exp(-sx)$ function. The L for the equation (2.2) is $L = \dfrac{\ln(2)}{s}$, which will be the reference we are going to use for calibrating s_1 and s_2. Every s is computed as $Ln(2) = \exp(\mu + k\sigma)$ with $k = \{-3, -2.5, -2, \ldots, 2.5, 3\}$, i.e., we compute ESL parameters to adjust the exponential function by varying the limit in 0,5 σ each for the Log-normal variable. Due to s_1 and s_2 must be different each other, the total number of combinations reach $\binom{13}{2} = 78$. From this, the best pairs of k to calculate S used is the one that obtains the best adjustment performance when applied in (2.1):

K1	K2
-2.5	0.5
-2	1.5

Table 2.1: Values of K to calculate the S value for the MGF

With this, we get a system of equations, in which we know the distribution of every single edge and their correlation with the other edges, in particular with which compound the path. So in the case of log-normal distribution, with this two parameters known for every edge, in [6] the authors built a system of equations in which we can get the parameters of the single distribution of the path that we want (the left side of (2.1). This is an important thing, because in this way its possible computing the distribution of a single path, easier to work instead of the sum of the distribution, which is possible represent a lot of edge and therefore will be very difficult use to.

To be able to solve this system of equations its necessary get two things:

1. The distribution of the sum of random variables that compound the right side of (2.1)
2. Known which distribution use to approximate the travel time distribution of the path

In the work [6], he study and determine that use that a a sum of log-normals follow a log-normal distribution approximately. The last important thing is that to compute the convolution of the sum of random variable, that is, the right side of (2.1) is a better approach to use a quadrature according to the structure of the argument of the integrals, in the case of log-normal an useful approach is the Gauss-Hermite. Use this quadrature series and then calculate the value by use a Montecarlo simulation through a method develop in [6] called "Equivalent sum of Log-normal method (ESL)", described in [9].

The ESl method let use to get an accurate representation of the MGF of the right side of [2.1] using samples of the distribution of the sum of random log-normals and through that compute the integrals of the convolution. From the joint distribution of the sum of distribution mentioned above, we take a sample of a Normal random variable X of length M, then we compute the mean of the exponential product over the sample, with this we have a sample $Z = (z_{11}, z_2, \ldots, z_M)$ of Normal random variables. This sample, let us compute the media of $\prod_{k=1}^{K} \exp\left(-s \cdot \exp\left(z_{\mathrm{i}k}\right)\right) for \mathrm{i} = 1 ... M$. In this way we have the montecarlo estimation of our MGF.

$$\widehat{\Psi}_{\left(\sum_{k=1}^{k} Y_k\right)}\left(S_{\mathrm{i}}, \mu, C\right) = \frac{1}{M} \sum\nolimits_{\mathrm{i}=1}^{M} \prod\nolimits_{K}^{k=1} \exp\left(-s \cdot \exp\left(z_{\mathrm{i}k}\right)\right) = \frac{1}{M} \sum\nolimits_{\mathrm{i}=1}^{M} g\left(Z_{\mathrm{i}}\right)$$

The motivation to develop this method is that it is difficult to compute the convolution especially if we are adding a large amount of random variables.

In summary, to apply this method is necessary consider:

1. First, distribution of every edge, following the same distribution
2. Second, distribution to approximate the sum of the distribution of the edges, for example the sum of log-normal fit well if we consider a log-normal to approximate
3. Third, the distribution of the sum of random variables, considering their correlations
4. Fourth, A quadrature to approximate the convolutions for computing

This four needs are the limitation in [6], if he doesn't have any of this points the method it cannot be executed. So, in this work we try to generalizes this previous work to allow the consideration of other distributions on the edges and path

2.4.2 Data in the Previous Work

To work with real data, Felipe Lagos took a GPS position of 11 different services routes of Transantiago (the public buses service in Santiago) that register their position every 30 seconds, this includes latitude, longitude, date and time and the information about the bus service such as vehicle registration plate, bus route, way that bus follow and instantaneous speed.

The data that he manage come from a complete month of information of the system, June of 2010, between 7 am to 11 pm. Once we have the data, the challenge is to project this GPS information into a graph of Santiago. For this, we have a directed graph representation of the city, more than 650,000 thousands of arcs and 350,000 nodes.

The projection of the information into the graph, mentioned above is made by an approach develop by [5]. This algorithm interpolates the projected points in segments to make them comparable each other. In our case this means that the GPS projections are interpolated to the extreme of the arcs, so we could estimate the real speed and time easily for the whole path.

To summarize, the algorithm of projection found what is the most logical path that follow a bus to determine the time that takes to get along a set of edges between two GPS points, from that it can be calculated the travel time. It's important use an algorithm because is possible project in a wrong combination of edges and subsequently calculate in a wrong way the travel time in the graph. Here we want to highlight the projection that is made, however it has a more complicated technical background than what is mentioned here, which is fully developed in [6]. The process of the projection it shows at the Figure 1.

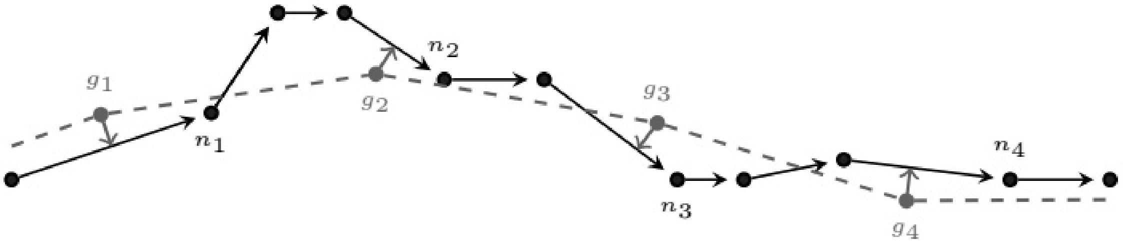

Figure 2.1: GPS example points of a bus projected into a map

Here the points of GPS is projected over the graph, where one point of GPS could be more than one node away from another, then it is assumed that the vehicle take the path underlying. In the Figure 2 it represented one path over Santiago map, in which the GPS is projected.

The travel time distribution of the variables are calculated from the data, between 7 am to 11 pm, the range in which are more data to estimate a good distribution for this variables. This random variables are included in 4876 edges and of Santiago's graph using 1026 buses that traveled along those edges.

This times are associated to 4876 different arcs of Santiago's graph, which are produced by 1026 buses that traveled along those arcs during June of 2010. All this data is stored in our database, what lets us to study the nature of travel time distributions

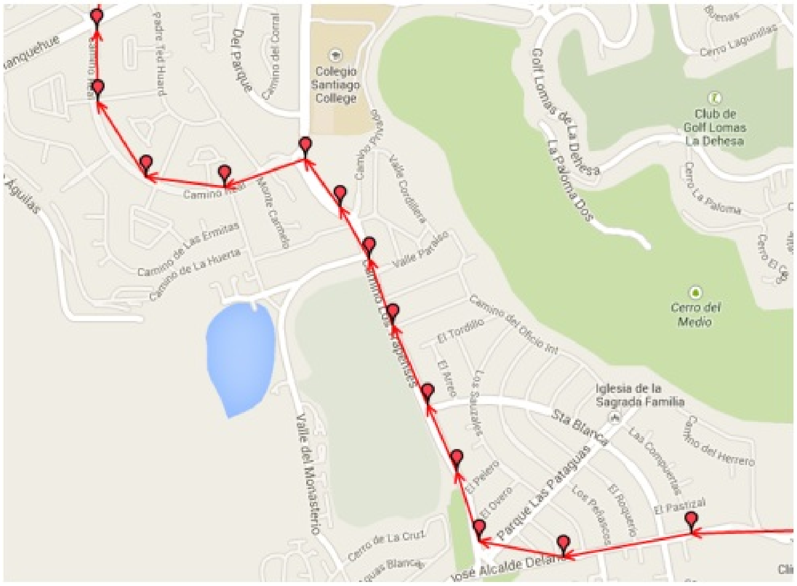

Figure 2.2: GPS example points of a bus and the projection of them in an edge set

3 A Simulation Approach for the Sum of Generic Random Variables

3.1 Methodology

The original methodology, described in the chapter 2.4 was previously described, in summary:

1. Describe the edges of a graph by a travel time distribution. It's important that each edge of the graph must have the same probability distribution. Also is necessary get the correlations between the edges of the graph
2. Described the paths in the graph as a single distribution. This distribution need to fit well as the sum of every random variable that compound the path, that is, the convolution of the random variables and his correlations
3. Get the joint distribution of the sum of random variables that compound the right side of (2.1)
4. Get a way to numerically approximate the resulting integral of the random variable that represent the path, the left side of (2.1). This way is probably a quadrature, like the one used by the original method
5. Sampling the joint distribution to apply an analogous use of ESL described by [6] over another distribution
6. Get the "S" value of the FGM to get the adequate number of equations to solve the system and get the parameters
7. Apply the formula (2.1) but in this context, that is, solve the non-linear system of the MGF of the sum of joint correlated distribution and the single distribution that approximate it.

Al this steps are necessary to approximate the distribution of a path from the edges that compound it. In this way is that this work develop certain owns methodologies that complement the steps above.

To begin, to get the distributions of the edges and the correlations is simple because it depends about the structure of the graph by itself, there is no way to modified a real structure. This is specially important when is working over a real data but if the work is in a controlled environment the things are kind of different.

In this work, for academic purpose in the first part, is not necessary create a complete graph it's only necessary simulate different set of paths with there respectively edges that conformed it, the final number of edges of this paths are set randomly by an uniform distribution between 3 and 15, that is, the paths has minimum 3 edges or maximum 15. The minimum is set because less than 3 is meaningless about a real necessity for approximate the distribution of the paths, and the maximum because with 15 we can see the effect over several edges compound a path but not in a extremely complex way that could causes over-demand computing processing. Also, more edges that 15 suggest that perhaps a better approach would be based on the Central Limit Theorem, although the theorem treats independent random variables, it would be an approximation to consider. Also, to get the correlations between the edges, they were sampled by a Normal(0,1).

Because it is necessary in the step 3 to know the joint distribution with correlations of the edges is that the chosen distribution to model the edges have to have a known joint distribution, if not then we cannot continue with the approximation, that is because with a known joint distribution the MGF could be developed and the simulations corresponding to the model performed. This is an important point for the other steps, because conditions the simulations and from there on everything else.

To get the distributions of the paths in the graph, that is, the distribution target to approximate the sum of the correlated edges is that we what was mentioned above, once established which are the distributions and correlations of the edges and the joint distribution between them, sampling the joint distribution to get values that follows the distribution of the paths, that is, the sum of the correlated distribution. The sample of the joint distribution will be done using the Acceptance-Rejection method, the detail of how develop this method is going to be described in the next sections.

Once we get the simulated data we need to get the final distribution of the path, that is, from the values simulated to the distribution representation, we propose to check several distributions and how they fit to the curve of the data to get the best approximation between this candidates, the fit of the distributions will be calculated maximizing the likelihood estimate (MLE) but to avoid the problem that the maximum likelihood found may be maximum locally is that the fit to the simulated data will be made 200 times, changing the initial parameters depending of the distribution to get several approximations from different starting points. This will be done 1.000 times in order to get different samples and then different fittings to the curve for each distribution. In each of this 1.000 iterations will be created a ranking of the distribution's fit, from the best to the worst (following the maximum MLE). Finally, from this ranking we could get the frequency of each distribution on each place of the ranking, selecting not necessarily the most frequent distribution in the first place, but the distribution that will have a better performance in general.

This way to get the distribution from the edges let us to calculate the distribution that fit well with the edges that compound the paths independent of the distributions of the edges because this part of the methodology can be calculated every time it needed.

It is also important to mention how we choose the way to numerically approximate the integral of the left side of (2.1), this integral is the result of the MGF of the path's distribution. To calculate this integral we are going to use a quadrature that depend on the distribution of the paths, this is essentially because the quadrature works over certain structure of the integral, so the chosen distribution will required some work to transform and in that way fit to some quadrature that we choose at the correct moment.

We now need to keep in mind all this steps and the way described above to solve, because they will be necessary when the distribution for the edges become chosen and the methodology will be develop to fit the distribution mentioned. Now the challenge it's apply this into a real situation, first with simulated data and then with real data.

3.2 Validation

In order to verify if this methodology has good results, a test was carried out using lognormal for the final distribution of the road and a joint lognormal distribution for the arcs that make it up, that is, the methodology was developed using the same basis as the previous work for to be able to have a reliable point of comparison, since it is known that these distributions work, and decide if this extension is being developed in a good way. As the S value of the FGM parameter, the same as in the previous work was used, for the same reason. Using all this elements we work in validate if fitting the distribution in the way mentioned above it's possible to approximate the result of the single distribution that approximate the joint distribution with correlations.

This thing is in particular important because if we can approximate the distribution in this way, then implies that we can use this to get, from the joint distribution, which is the distribution that fit in a better way. From here we released one of the restrictions that the original methodology had, this is what is the distribution that approximates the sum of random variables that make up the edges?

Remembering that the way in which we fit the distribution to get the parameters of the distribution that approximate the joint distribution may fall into local maximum we sampling once and fit the distribution 200 times, changing the parameters following a uniform from 1 to 1000. Then we keep the parameters that have a better MLE from the estimation. This was done 200 times, to get a complete result of this way to approximate the joint distribution . To get an idea of each result of this 200 execution in the next figure we observe an histogram consisting of a sampling following a sum distribution of correlated lognormal, the approximation in this case of the single lognormal that approximate and the better fit function found by the fit method that we use and was mentioned before:

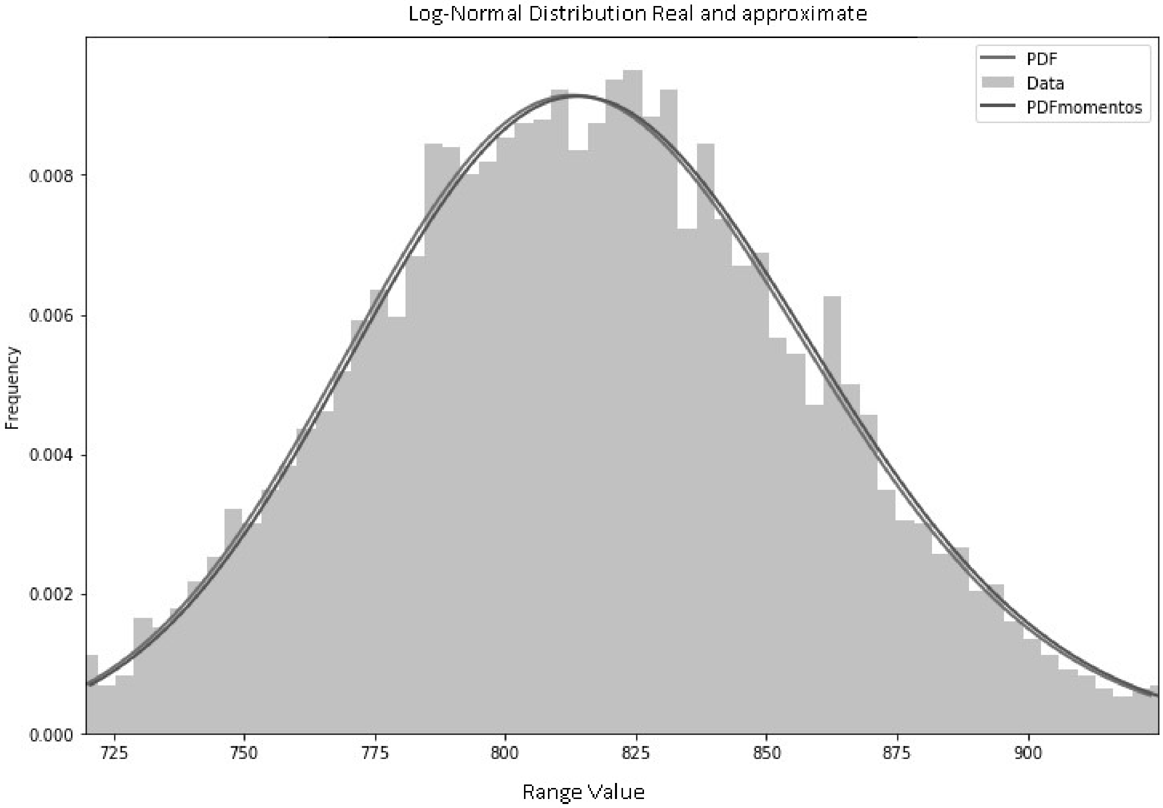

Figure 3.1: Lognormal-FGM

The curve corresponding to “PDF” has the form of the lognormal distribution to which the sum of the correlated random variables would approximate according to the parameters calculated by the .fit method of the program described above. The "PDFmoments" curve corresponds to a lognormal distribution with the parameters calculated by the method of the moments developed here. The legend "Data" is the simulated data. Visually there is a great similarity between the calculated and simulated distributions, which is positive when evaluating what we propose about this methodology. Once the parameters have been estimated, globally in all the executions of the algorithm, it is observed that there is a difference at the level of the mean calculated of 2% and at the level of standard deviation of less than 0.3% which meant an increase in the standard error of 1%, it is considered adequate to continue.

Parameters	Average Error
Mean	2%
Standard Deviation	0.3%
MLE	1%

Table 3.1: Parameters and their average error against simulated data

3.3 Distribution to Work

3.3.1 Joint Distribution to Work

To apply this methodology in a simulated environment we need to solve an apply all the steps described in the section 3.1. It is important to prioritize which are the most important needs in order to implement all the steps, so as not to make decisions that in the future represent an obstacle when developing this algorithm.

As described before the development of the methodology implies, in different steps, to simulate the joint distribution of the edges and that's why define the joint distribution of the edges is the priority, because from it everything else is developed.

Doing a little research of joint distributions with definite correlations is that we find the work done by Nadarajah [11] in which he made a review of some results on sums of random variables, in this way is that we choose the gamma distribution as the distribution of the edges that we simulate, mainly because this is the only one that it's worked using correlations between these random variables.
The sum of gamma with correlations is model as follows in the review.

Let $Y_{\mathrm{i}} = 1, 2, \ldots, N$ be dependent gamma random variables with parameters $(\alpha, 1/\beta_{\mathrm{i}})$ and $\rho(\mathrm{i}j)$ the correlation between Y_{i} and Y_j, $\mathrm{i} = 1, 2, \ldots, N$ and let $Z = Y_1 + Y_2 +, \cdots + Y_N$ The main thing here is the representation of Z. Alouini et al. [2] provide a generalization for the sum of gamma random variables when this random variables are correlated. Then the density of Z can be express as

$$f_z(z) = \prod_{n=1}^{N} (\frac{\lambda_1}{\lambda_n})^n \sum_{k=0}^{\infty} \frac{\delta_k Z^{N\alpha+k-1} \exp\left(\frac{-z}{\lambda_1}\right)}{\lambda_1^{N\alpha+k} \Gamma(N\alpha + k)} \tag{3.1}$$

for $z > 0$, where $\lambda_1 = \min \lambda_n$, λ_n are the eigenvalues of the matrix A=DC, where D is the N x N diagonal matrix with the entries β_n and C is the N x N positive definite matrix, because it corresponds to a symmetric matrix, define by

$$C = \begin{bmatrix} 1 & \sqrt{\rho_{12}} & \cdots & \sqrt{\rho_{1N}} \\ \sqrt{\rho_{21}} & 1 & \ldots & \sqrt{\rho_{2N}} \\ \vdots & \vdots & \ldots & \vdots \\ \sqrt{\rho_{N1}} & \ldots & \ldots & 1 \end{bmatrix}$$

and the coefficients δ_k satisfy the recurrence relations

$$\delta_0 = 1$$

and

$$\delta_{k+1} = \frac{\alpha}{k+1} \sum_{\mathrm{i}=1}^{k+1} \left[\sum_{j=1}^{N} \left(1 - \frac{\lambda_1}{\lambda_j}\right)^{\mathrm{i}} \right] \delta_{k+1-\mathrm{i}} \tag{3.2}$$

Here it is important to note that an analogous way of modeling this joint distribution corresponds to building, from the gamma $(\alpha, \beta_\mathrm{i})$ with a correlation matrix C, gamma distributions with λ_i instead of β_i, independent distribution, where λ_i is indicate above how to calculate. Although the second form seems more approachable, the first form is used in this instance and later this second form will be used.

3.3.2 Approximate Joint Distribution

Once we define the gamma distribution as the one that explains the behavior of the edges, it's that we must to develop the point 2 of the steps in the section 3.1, that is, find the distribution that approximates the joint distribution of the edges.

As we described before, we are going to generate a different samples that follows the multivariate gamma with correlations that we described before in this section. To generate this samples we first get the set of gamma distributions, this was obtained in sampling N random times the β parameter of each distribution and one the parameter α for the set of gamma distribution and the correlation matrix for this N variables, in specific:

1. The number N is sample between 3 and 15 with a uniform distribution, this was mentioned in the section 3.1
2. The β parameter was sample following a uniform between 1 to 15, these numbers were chosen because we tried to reach the highest range of values that the computer can manage in a reasonable time
3. The parameter α was sample following a uniform between 1 to 20, this numbers were chosen for the same reason that the before point
4. The correlation matrix was sample following a N $(0, 1)$

With this we have fully defined how get f(x) and g(x).

Now, to sample this multivariate distribution we need to work with the formula (7.1) about the joint density, but it's a complicate formula, so as we mentioned in the chapter 6, we are going to use the Acceptance-Rejection method that we described in detail in the chapter 2, section 1. But to use this method of sampling we need two things:

1. A distribution g(x) that is "close" to the multivariate gamma we want to sample f(x), our distribution , following the nomenclature used in background
2. The parameter "c", that bounded the ratio f(x)/g(x)

First as g(x) it's proposed to used the distribution gamma but without correlation, that is, a variables gamma $Y_\mathrm{i} = 1, 2, \ldots, N$ with parameters $(\alpha, 1/\beta_\mathrm{i})$ and $\rho_{\mathrm{i}j} = 0$; $\forall_{\mathrm{i},j}$, the implication of this is that $\lambda_{\mathrm{i}} = \beta_{\mathrm{i}}\ \forall_{\mathrm{i}}$. Second, as "c" we calculate, every time the set of gamma distributions change, the minimum of the function h(x)=f(x)/g(X) in 500 different starting points equispaced between 1 to 10.000, this to reduce the probability to fall in local minimums. It's chosen as "c" the minimum value founded of h(x).

Here is important to mention that about the correlation matrix simulated for the multivariate gamma with correlation, it's gonna be used only when we consider the correlated gamma variables, but when we use the distribution g(x), that is, the multivariate independent gamma, this matrix it will be considered as an identity matrix, for logical reasons.

Now, having defined the necessary elements to develop the sampling, that is, g (x), the value "c" described above, the joint distribution with correlations of the edges and a way of writing it's that the samples are calculated. To perform the sampling of f(x), the Acceptance-Rejection method is performed on a sampling basis of g(x) of 10,000 data, from which the data are obtained with the desired distribution. Once these data are obtained, after applying the sampling method, the adjustment to the curve of a set of distributions is calculated, these distributions are the candidates to approximate our multivariate correlated gamma distribution. The adjustment is made using the method of MLE, as mentioned in the previous chapter, it should be noted that for each distribution the adjustment was made 200 times in order to avoid falling into local minimums and considering the best adjustment as the one where the likelihood is maximized. The distributions used to adjust and propose as candidates for the best fit are: Normal, Exponential, Uniform, Erlang, Gamma, Johnsons-Su, Johnsons-Sb, Weibull, Gumbel, Burr, Nakagami, Rayleigh, Rice and Lognorm.

This methodology was performed 1.000 times obtaining each time:

1. A set of multivariate correlated gamma
2. A set of data that follows the desire distribution
3. The best adjustment of each distribution to the data sampled

It should be noted that we defined as g(x) the multivariate independent gamma but is not necessarily the best choice, although in this context is enough.

3.3.3 Results of the Approximation

After iterating 1.000 times is created the a ranking from the results of every iteration. This ranking corresponds to the frequency of the performance of each distribution in each iteration, in other words, every iteration each distribution is ranked from the best fit to the worst, that is, the distribution with the maximum likelihood is, in this context, the best and the distribution with the minimum likelihood is the worst, then we calculate the frequency of each iteration to get the final ranking. This ranking is important to analyze it not only which one has the bets performance in frequency, also is important the distribution of its results in the ranking, in other words, if the distribution with the best ranking, when it is not the best is the worst, it may not be good to use it, since there is a percentage of the times when using it we are mistakenly modeling the joint distribution of the edges, in this case the multivariate correlated gamma.

After performing the iterations and building the performance ranking, five distributions obtain between 76% and 86% of the first 5 places in the ranking, so it is considered that among these 5 distributions the best results are obtained. These results are shown in the table below.:

Distribution	Performance Ranking				
	First	Second	Third	Fourth	Fifth
Johnson-Su	31%	9%	5%	6%	24%
Johnson-Sb	26%	15%	16%	14%	7%
Gamma	13%	10%	30%	27%	14%
Lognorm	10%	25%	4%	7%	24%
Erlang	4%	17%	30%	32%	9%

Table 3.2: Ranking of best distribution fit

To understand this ranking we consider that the percentage is built on base of how many times of the 1.000 iteration the distribution is the nth best distribution. For example Lognorm in First has 10%, that is, the 10% of the samples, Lognorm has the best performance. This table was created in this way because it is a way of seeing "normalized" performance, since by adding the maximum likelihood, it is affected by the relative magnitudes of each simulation, so that the particular effects on the generality could be lost, in this way the ranking is important not to consider the best performance only, but also to observe the position of the best performances. The Johnson-Su distribution is the one with the best performance in percentage, but almost 0.75 of the time that is the best distribution is the worst, so it's a risky distribution to chosen. In addition, both distribution Gamma and Erlang accumulate more than 90% of its performance in this 5 places, then considering that the Gamma distribution has a better performance in terms of the places it reaches in the ranking, this distribution is chosen to approximate the multivariate correlated gamma distribution.

It is also important to mention that the distributions Johnson (Su and Sb) have four parameters unlike the others that have two, this is relevant to the extent that it is considered that with a greater number of parameters the distribution has more degrees of freedom, which allows it to better adjust to the curves. So if this situation is considered to penalize the adjustment of these distributions given the greater degree of freedom, the choice of the gamma distribution would be even more robust as the one approximating the chosen joint distribution of the edges.

3.4 Approaching Joint Distribution

Now that we have the function of joint density for the edges and the distribution to which we are going to approximate it we need to solve the next equation:

$$\Psi_y\left(S_{\mathrm{i}}, \alpha_x, \beta_x\right) = \Psi_{\left(\sum_{k=1}^{k} Y_k\right)}\left(S_{\mathrm{i}}, \alpha, \beta, C\right) \text{ for i=1,2 [1]} \tag{3.3}$$

This equation is the analogous to the equation (3.3), now we continue to establish 3 main elements:

1. How we are going to approximate numerically the integral of the left side of (3.3)
2. Apply the analogous form of the ESL method to the multivariate correlated gamma distribution
3. Calculate the values of S, the parameter of the MGF. In this work we used two values, because the gamma function have 2 parameters α and β

3.4.1 Numerically Calculate the MGF of the Target Gamma distribution

To calculate the left side of (3.3), we can develop two ways, the first is to directly use the closed-form that has the function generating moments of the gamma distribution, which is represented as follows:

$$\phi_y(s, \beta, \alpha) = (1 - \frac{t}{\beta})^{-\alpha} \tag{3.4}$$

The second form corresponds to propose the canonical form of MGF and approximate it numerically, in this case to continue the form applied in the previous work [6] we will use the canonical form.

So the canonical form of MGF:

$$
\begin{aligned}
\phi_y(s,\beta,\alpha) &= \int_0^{+\infty} \exp{(-sy)} f(y) \mathrm{d}y \\
&= \int_0^{+\infty} \exp{(-sy)} \frac{\beta^\alpha y^{\alpha-1}\exp{(-\beta y)}}{\Gamma(\alpha)} \mathrm{d}y \\
&= \int_0^{+\infty} \exp\left(\frac{-sz}{\beta}\right) \frac{\beta^\alpha \left(\frac{z}{\beta}\right)^{\alpha-1} \exp\left(-\beta\frac{z}{\beta}\right)}{\Gamma(\alpha)} \frac{\mathrm{d}z}{\beta} \\
&= \int_0^{+\infty} \exp{(-z)} \frac{\exp\left(\frac{-sz}{\beta}\right) z^{\alpha-1}}{\Gamma(\alpha)} \mathrm{d}z \\
&= \sum_{\mathrm{i}=1}^{N} w_\mathrm{i} \frac{\exp\left(\frac{-sx_\mathrm{i}}{\beta}\right) x_\mathrm{i}^{\alpha-1}}{\Gamma(\alpha)} + E_n
\end{aligned}
\tag{3.5}
$$

Here we are using the Gauss-Laguerre quadrature. This let us to have an equation that depends only by the parameters that we want to obtain. The coefficients w_i and x_i are known and can be obtained from [1]. N is the Laguerre integration order and E_n is a remainder term that decreases as N increases. Authors suggest that N = 12 is accuracy enough to get a good approximation of the integral so from this N E_n is not relevant term. So with this formula and quadrature we have a numerically approximation of the left side of (3.3). Here we use as α and β the parameters calculate when a sample of the joint distribution of the edges underlying of path is adjust by a gamma distribution using MLE method.

3.4.2 Apply an Analogous ESL

Following the same logic as the work developed by [6], obtaining the right side of the equation (3.3) in the same way as the left side, that mean calculate this quadrature for the convolution from the multivariate correlated distribution function, however this it's a complicated job to perform and computationally difficult to handle. For example, a path with 15 arches implies managing at least 15^2 coefficients, many of them, representing correlations, very small, so storing and operating them is not the most efficient. It's for them that here it is suggested to apply in the same way as for lognormals variables the ESL method suggested in the work of[6] nut now with gamma distribution.

The ESL method is to get an accurate representation of MGF using samples from the joined distribution of the gamma multivariate gamma and compute the integral. From joined distribution we take a M length sample of Y, our gamma variable, and compute the mean of exponential product part over the sample. In other words, this sample matrix $Z = (Z_1, Z_2, \ldots, Z_M)$ let us to compute the media of $\exp{(-s \cdot Z_\mathrm{i})} for \mathrm{i} = 1, \ldots, M$. In this way we have the Monte Carlo estimation of our MGF:

$$\hat{\Psi}_{\sum Y_k}(s,\beta,\alpha,C) = \frac{1}{M}\sum_{\mathrm{i}=1}^{M}\exp\left(-s\cdot Z_{\mathrm{i}}\right) = \frac{1}{M}\sum_{\mathrm{i}=1}^{M} g(Z_{\mathrm{i}}) \tag{3.6}$$

3.4.3 Calculate the S Parameter of the MGF

Once we have the equation (3.3) completely defined, it's necessary found the values S of the MGF to get the equation system from which we are going to found the parameters of the distribution that we are approximating. Following the same logic that has described in the previous work we consider that the MGF is in certain point of view a weighted sum of $\exp(-sX)$, where X is the random variable that follows, in this context, the gamma distribution. Therefore, to avoid moving this weighted sum towards the distribution tails and having a more adequate value, the following is done.
The mass center of the distribution L is :

$$\begin{aligned}\frac{1}{2s} &= \int_0^L \exp(-sX)\mathrm{d}x \\ &= \frac{1-\exp{-sL}}{s}\end{aligned} \tag{3.7}$$

here we use $1/2s = I/2$ because is the half of the total area for the integral for $\exp(-sX)$ function. The L for the equation [7.6] is $L = \dfrac{ln(2)}{s}$, which will be the reference that we are going to use for calibrating s_1 and s_2.

In order to calculate L, the values that this parameter can take are discretized and each combination is evaluated. This results in that s_1 and s_2, the two values of "s" that the MGF it's going to use, can take 78 different values, considering a certain range. Considering $L = \exp(\mu + k\cdot\sigma)$, where $k = \{-3, -2.5, -2, \ldots, 2.5, 3\}$, whereby $s = \frac{ln(2)}{exp(\mu+k\cdot\sigma)}$, where k is a factor μ and σ are those that correspond to the gamma distribution that is being adjusted, this means that the value "s" is calculated by varying in 0.5σ each time according to the gamma distribution, from the above that the number of combinations of s1 and s2 is $\binom{13}{2} = 78$.

However, to calculate the value of S it is necessary to haveμ and σ, distribution parameters that have not yet been estimated, therefore these parameters are calculated from the α and β parameters f the distribution using the MLE method mentioned above, applied in 78 groups of 60 samples of original size 10,000 on which the Acceptance-Rejection method is applied so that it follows the joint distribution of the sum of the correlated gammas random variables. This means that for each of the 60 samples, its parameters are estimated with MLE, then these will be a first approximation to use to calculate the S parameter.

Once you have the initial μ and σ, or what is the same α and β, to estimate a first value of "S", the pairs of s1 and s2 and the 60 independent samples, proceed to calculate the finalα and β. This calculation is made from the ESL analog method, this calculation is made for each of the 78 groups and in each group over the 60 samples. Each of these 60 samples comes from a set of different gammas random variables and a different correlation matrix as well. Having in each of the 78 groups the same 60 data distributions. Therefore, for each of the 78 pairs of s1 and s2 the adjustment with the ESL method will be 60 times different

Once the parameters in each sampling have been calculated, a Kolmogrov-Smirnov 2-sample test is carried out, in which the original sampling is compared, from which the estimate comes, against a sample obtained from a gamma distribution with the α and β parameters obtained previously. This sampling is the same size as the sampling against which it is compared, obtaining as a final result the average number of times that the sampling, based on the calculated parameters, is accepted and from there know, what are the K that allowed to build the S that get the best performance α and β parameters.

In the next figure we observe the performance of the average acceptance of the approximation gamma distribution for each of the pairs of s_1 and s_2. The main thing about this figure is to observe the range between the acceptance varies, in other words, here the performance of each pair of the MGF parameters is ordered from best to worst, so in the first places are those parameters where the goodness of fit test is not rejected the null hypothesis that the distributions come from the same distribution in a higher percentage of experiments. Then the results will be shown for when the acceptance is high.

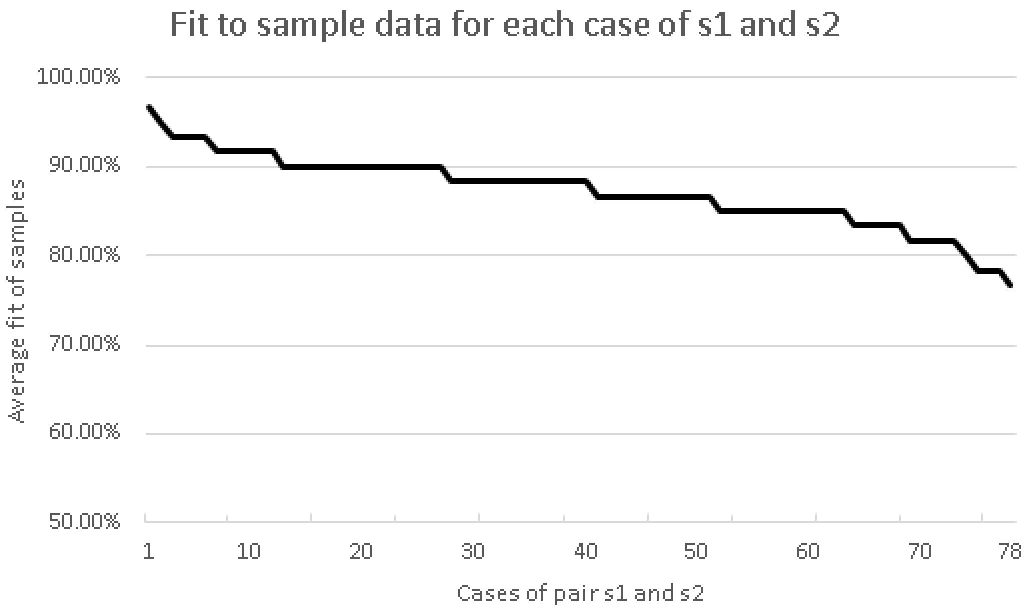

Figure 3.2: Acceptance of S

Here we observe that the range about the acceptance is between 76,67% and 96,67%, this means that between 76,67% and 96,67% of the cases the test indicates that it cannot be rejected that the samples come from the same distribution . These results are very positive, however it is limited to the 60 samples obtained which could not be increased due to limitations of computational capacity and available processing time, despite this the order of magnitude of acceptance should be the one obtained.

From the above, the results with a performance over 90% are 12, in each of which there are 2 values for K, which generate the values s_1 and s_2 that allow to complete the system of non-linear equations necessary to solve the equation [7.3], below is the table in which are presented the two best combinations of K values are presented for which s_1 and s_2 allow to obtain the α and β values that obtain the best acceptance results.

Acceptance percentage	K1	K2
96,67%	-0.5	1.5
95.00%	-2	2

Table 3.3: Better results for S according to K value

From here it is concluded that to construct the necessary system of equations, k_1 and k_2 will be used to construct s_1 and s_2, the values -0.5 and 1.5 respectively.

3.4.4 Apply the Joint Distribution Approximation

At this point we have all the necessary elements to solve the system of equations raised in the equation (3.3), that is:

1. The quadrature to calculate the left side of (3.3)
2. The analogous ESL to calculate the right side of (3.3)
3. The s_1 and s_2 necessary to get the system of equations to calculate the parameters of the distribution on the left side of (3.3)

Given the above is that the estimation of the parameters was made, but in order to have an integral a comprehensive understanding of the effect of this methodology is that the calculation for 80 different configurations is carried out, these were elaborated by varying the α value of the distribution between 1 and 20, choosing 10 values in a balanced manner and varying the size of the sum of random variables between 3 and 10, in this case the amount of variables to be added was reduced, mainly due to a practical issue and to be able to more easily observe the marginal movements in the configuration changes.

In summary the method was applied as follows:

1. To execute the methodology a high number of times, independently and thus observe its performance is that 160 different distributions are constructed for each configuration of alpha and size, this implies that a total of $(80{\cdot}160 = 12.800)$12.800 samples is developed independent. Each of these samples is developed using the Acceptance-Rejection method from a base of 10.000 distribution data that approximates our joint gammas distribution
2. For a fixed size and alpha (those shown in the table below), the parameters β of the 160 distributions that will be used to test the performance of the method in each of the configurations to be evaluated are chosen in a uniform random manner between 1 and 16
3. The parameters that approximate the joint distribution for each of these samples are calculated with the resolution of the equation (3.3)
4. A sampling is calculated for each set of calculated parameters, this is, for each configuration and for each sample of the 160 used, we calculate a new sample (sample B) with the parameters that were previously calculated and with the same size of the sample from which it was calculated (sample A)
5. It's calculated the average number of times , by configuration (that is, 160 times by configuration), that when applying the Kolmogrov-Smirnov test the hypothesis that the original sampling (sampling A) and the one calculated from The parameters calculated when solving the system of equation (3.3) (sampling B) come from the same distribution

From the steps described above, we get the next table. This table shows the result mentioned when applying the goodness of fit test on the samples developed by the methodology, comparing the "real sample" with the distribution suggested by the approximation.

		Size (Number of sum of RV)							
		3	4	5	6	7	8	9	10
α	1	72.96%	69.81%	74.21%	85.00%	84.38%	82.50%	80.00%	83.13%
	3.11	81.25%	86.88%	85.00%	85.00%	86.25%	90.63%	85.63%	91.88%
	5.22	86.71%	90.63%	87.50%	89.38%	88.75%	87.50%	86.88%	90.00%
	7.33	89.24%	88.13%	91.25%	91.88%	88.75%	91.88%	91.25%	88.13%
	9.44	88.68%	88.68%	87.50%	90.63%	93.13%	90.00%	92.50%	87.50%
	11.56	88.13%	86.88%	86.25%	87.50%	87.50%	88.13%	88.13%	94.38%
	13.67	84.38%	90.00%	91.25%	91.25%	89.38%	88.75%	86.88%	87.50%
	15.78	89.38%	90.63%	87.50%	88.75%	91.88%	90.00%	88.13%	93.13%
	17.89	90.63%	90.63%	88.75%	88.75%	89.38%	85.63%	97.50%	96.88%
	20	88.13%	91.88%	91.25%	90.00%	93.75%	96.88%	98.75%	99.38%

Table 3.4: Average result of the approval of the Kolmogrov-Smirnov goodness of fit test of the sum of random gamma variables, adjusted by a gamma distribution, by different size and shape parameter α

This table shows first that acceptance in general is high, above 70%. In addition, it is observed that as the α value is higher, the test tends to pass, on average, more. The same happens when the size of random variables involved increases. The first may be due to the fact that by increasing the value of alpha, the distribution tends to behave more like a normal distribution, so it is easier to calculate the behavior that a "normal" multivariate function would have, this would be explain by the relation of the skewness with the shape parameter by $Skewness = 2/\sqrt{\alpha}$. The second may be because there are more edges, the approach into a single distribution that this method propose is more natural and avoid to give so much weight to particular behaviors of some distribution in particular.

From what has been said here, it would be interesting to evaluate the performance of a normal distribution to approximate the multivariate gamma distribution, which although it should not give better results than those obtained here since it was verified when choosing the distribution to which this multivariate distribution approximates, that the best performer was the gamma, even above normal. That said, the same exercise was carried out approximating the normal distribution to the multivariate gamma distribution, for this the k values described above that define the s values corresponding to the MGF were recalibrated. Once the parameters have been calculated, the method is applied as described above but this time using a normal distribution, calculated numerically by means of the Gauss-Hermite quadrature, similar to what was done in [6], to approximate the gamma distribution multivariate. The results are summarized below.

Acceptance percentage	K1	K2
14.38%	-1	2.5
12.5%	2	3

Table 3.5: Better results for S according to K value for normal distribution

The table above summarizes the two best results of the k values that allow reaching the best s parameters for the MGF that is used with the normal distribution. These results are far from those achieved by calibrating the parameters for normal distribution. These parameters are used to calculate the results of the normal distribution below.

		Size (Number of sum of RV)							
		5	10	15	20	25	30	35	40
α	1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	3.11	0.00%	0.00%	1.25%	1.88%	0.63%	3.13%	3.13%	7.50%
	5.22	0.00%	3.13%	4.38%	11.25%	11.88%	11.88%	20.00%	25.63%
	7.33	3.75%	6.88%	10.63%	11.88%	21.88%	27.50%	28.75%	37.50%
	9.44	5.63%	12.50%	18.13%	28.13%	30.63%	33.13%	43.13%	46.25%
	11.56	8.75%	16.25%	22.50%	36.25%	34.38%	43.13%	48.13%	50.63%
	13.67	15.00%	26.88%	28.13%	41.88%	45.63%	48.13%	51.88%	55.63%
	15.78	16.88%	3.13%	41.25%	44.38%	52.50%	61.25%	55.63%	60.00%
	17.89	24.38%	36.88%	30.00%	41.25%	46.25%	56.25%	64.38%	63.13%
	20	26.25%	36.25%	46.88%	46.88%	60.00%	67.50%	68.13%	66.25%

Table 3.6: Average result of the approval of the Kolmogrov-Smirnov goodness of fit test of the sum of random gamma variables, adjusted by a normal distribution, by different size and shape parameter α

Here we observe the behavior of adjusting with a normal distribution the joint distribution of the gamma random variables. It is mainly highlighted in the table that the average adjustment behavior, for a large set of paths, begins to be much better, this means, that when the number of paths that participate in the joint distribution increases, the adjustment behavior of a distribution normal is better. In particular for 35 arcs and an α value of 15.8,56% of the normal distribution is accepted according to the Kolmogrov-Smirnov goodness of fit test. This goes along the line of the central limit theorem mentioned above, which indicates that for a general case, with a large number of paths that participate in the joint distribution, it is possible to use a normal distribution, however this does not work properly when the number of paths is small, such as when there are 5 or 10 arcs.

It is also interesting to note that the normal distribution adjustment compartment improves when the α value increases, this is mainly due to the fact that an increase in the shape parameter decreases the skewness through the relationship $Skewness = 2/\sqrt{\alpha}$, as mentioned above, which tends to the gamma distribution, at the limit, having a skewness equal to 0, just like a normal distribution.

3.5 Testing the Method with Real Data

In the sections above we present a method to get the distribution of a path from the underlying edges and their correlation, but all the results and the application are applied with simulated data, this is a good approach to work and proof the concept that is presenting. But to get a complete idea of this work It is interesting to ask if this methodology is applicable if real data is used, so to perform a proof of concept of this point it is that a real data structure on which to apply this methodology will be used. In this context it is that the best approximations to the structure that was previously raised from the real data will be chosen, this in order to prove that in conditions such as those studied it is possible to apply the methodology, otherwise we would have to rethink the distributions gamma raised (because, maybe another structure of distributions or correlation is more accurate), however for this purpose it was previously stated how to proceed in case of having a different data set.

3.5.1 Data and Structure

To apply what was stated at the beginning of this section is that the actual data used in the previous work described in chapter 2 section 3 subsection 2 will be used, that is, the real data used and transformed in the work of [6].

In summary, we have a full month of information on the Transantiago bus system, from June 2010 generated by 11 services, these record their position and speed every 30 seconds, so to know the travel time just calculate the distance traveled in each GPS measurement. Although [6] considered a subset of the data, we will take the totality of the data and we will filter it to have the most similar data, in terms of structure, to which we have raised throughout the work (gamma distribution). These filters are described in detail below.

In this sense, we were given the information of the graph on which these GPS positions and the underlying speed are projected. The graph has a total of 662,743 arcs and 325,262, each edge is composed by a unique identifier index, an index that indicates the beginning node and another the end node of the edge, in addition the nodes are constituted by a unique identifier index and the latitude and longitude where they are located geographically. Also, we have the velocity information (more than 8.5 millions of data), which corresponds to a list that indicates, among other things, for each measurement the measured speed, a reference to which edge corresponds and the exact datetime when it was measured, that is, there is a set of measured velocities of the edges of the graph with the datetime at which they were measured. From here we proceed to the transformation of the data in the desired structure.

First, to obtain the travel times, the day was divided into 48 hours, that is, the day was represented by 48 half-hour blocks, this was done since during the day it is expected that the distribution of travel times changes, but In half-hour blocks it can be assumed that it is homogeneous enough. To get an idea of this point, we can calculate how many edges with enough data follow a gamma distribution applying the test Kolmogrov-Smirnov with the real data against the distribution gamma adjusted by the MLE method. Here we obtain that for each block of half hour the percentage of edges that they cannot reject the hypothesis that their real data comes from a gamma distribution fluctuates between 99% and 35%, that is, the hypothesis that the real data comes from a gamma distribution is stronger or less according to the chosen half hour , which makes clear the mentioned point considering that in each block there are the same edges and arcs, only changing the measured data that varies according to the behavior of the city in that time window. In other words, there are times in the day where the distribution resembles what is sought and other moments where it does not

Then proceed as follows to obtain a set of paths that follow a structure similar to that of previously worked. The analysis of what edges to use was performed for every half hour, this means that if the decision is made not to use an edges, then that edge will not be used in the half hour in which the decision was made, but it could be used in another block of half hour.

- For every half hour, all travel time measurements that were measured in that half hour were stored by edge to which they belong

- In every half hour, all edges that have less than 15 measurements were removed from the analysis, this is because otherwise there would be no representative sample of the travel times of that edge in that half hour

- It was constructed, from the information of edges and nodes, all the paths of the graph of size between 3 and 13, that is that the amount of edges that compose the paths are between those magnitudes

- Once the edges to be used in every half hour are filtered, all previously calculated paths that have at least one of their edges outside the analysis are removed from the analysis. This was done because path loses one of the elements that constitute it

- Then, to avoid over representing the information of certain edges (which can be positive or negative), all paths that are subset of another or that share some edge are eliminated from the set to be analyzed, that is, in a half hour, the paths are ordered by size from largest to smallest and paths that have no edge shared with another path that has been previously accepted are kept for analysis. It should be noted that since the sets chosen for each half hour are assumed differently, it may be that the edge of a path in a half hour appears in another path in another half hour, but not in the same one to which it belongs

- Subsequently, for each path that was included in the analysis we adjust a gamma distribution for their edges. The value of α and the different β of the distributions of the edges that make up the path that allow maximizing the average p-value of the distributions is calculated (this adjustment is made by MLE), this p-value is obtained for each distribution when the Kolmogrov-Smirnov goodness of fit test is applied to test if it can be rejected that the measured data and the gamma distribution with the adjusted parameters come from the same distribution or not. In other words, the best configuration of the alpha and beta values is found so that when applying the test on the adjustment of the distribution of each edge and its measured data, the average p-value is the maximum

- Once the β values and the fixed α value for each path are calculated, all those paths whose edges when the Kolmogrov-Smirnov test is applied, less than 30% is passed when comparing the parameters and the measured data are removed from the analysis. That is, those paths that after adjusting their edges to a gamma distribution, 30% or less of these edges do not adjust their measured data to this adjusted gamma distribution will not be considered. In summary, what is intended at this point is to choose a sample that would have passed the tests that would lead us to use the model already described above. Although the elements are chosen a posteriori, an environment is recreated with real data in which the gamma distribution would have been applied, all this as a proof of concept with real data

- Finally, the method developed in this work was applied with the remaining paths, which represent 735 in total, here no distinction is made by half-hour blocks, since after the filters applied all the chosen roads should be different, at least in as for distribution of travel times

Second, it should be noted that the filters made are to obtain a structure similar to that developed in the simulation stage, however the filters applied are at the edge level and not for the paths since it was previously established that the sum of gamma distributions can be approximate as a single gamma distribution, which should happen in this context if only sets of edges are chosen for the paths that follows a gamma distribution, every structure of the edges is a consequence of the underlying edges structure.

It is also important to note that although a 30% level is low when dealing with edges that, according to the Kolmogrov-Smirnov test, come correctly from a gamma distribution with the adjusted parameters, as detailed above, It makes sense to consider that it would be naive to believe that all edges will distribute gamma, so it is a more realistic situation and also if good results are obtained, maintaining this consideration, it would be possible to understand the deep scope of this methodology.

3.5.2 A Real Data Approach

In the previous subsection it was fully described how to obtain the actual data to be used to apply the method in real data. In this sense, the application of the method applied to the structure of the data obtained is described. As a reminder, the final application of the method corresponds to the resolution of the equation (3.3)

First, to solve the left side of the equation (3.3), proceed as described previously in subsection 3.4.1 by equation 3.5. This equation corresponds to the quadrature applied to the MGF integral of the gamma distribution that follows the travel time along the path. However, in this iteration the α and β parameters of the equation that will be used will be those resulting from adjusting a gamma distribution, using MLE, to the actual travel time values measured in that edge for the half-hour block in which it was selected .

Secondly, to approximate the right side of the equation (3.3) we proceed to apply the ESL method for a gamma distributions, however, for this point of the work the sample of this distributions it is carried out according to the second way describedaccording to the definition made in section 3.3.1, that is, a set of gamma distributions with α and β_{i} parameters (for this set of real data, the parameters were calculated in the previous subsection) with correlation matrix C (which, for this set of real data corresponds to the coefficient of pearson between the edges according to their behavior of measured data) and a variable $Z = sum_{\text{i}} Y_{\text{i}}$, which represents the joint distribution of gamma distributions, the variable Z that has a density function described in 3.1, has the same distribution as a variable $\hat{Z} = \sum_{\text{i}} \hat{Y_{\text{i}}}$ where $\hat{Y_{\text{i}}}$ follows a gamma distribution $(\alpha, \hat{\beta_{\text{i}}} = \lambda_{\text{i}})$ with correlation matrix C $=$ I, where I is the identity matrix and λ the eigenvalues of the matrix $A = DC$, where D is the diagonal matrix of the β_{i} parameters. That is, $\hat{Z}$ is the sum of independent gamma variables with α equal to the α value of the original distributions, and $\hat{\beta}$ parameters equal to the λ_{i} parameters specified in subsection

3.3.1, more details in the work of [2]. So, from this point, to apply the analogous method of ESL following the equation (3.6) we simulate M times independent gamma variables $\hat{Y}_{\mathrm{i}}$ with parameters $(\alpha, \hat{\beta}_{\mathrm{i}} = \lambda_{\mathrm{i}})$ and then $Z_m = \sum Y_{\mathrm{i}}$, $m = 1, \ldots, M$. Here we have M samples necessary to apply the equation 3.6 and the size of M is equal to the amount of measurements that are taken from the travel time in the path, this to be consistent with the amount of measurements that were used to estimate the parameters of the gamma distribution to which we are going to approximate (parameters which were calculated in the first point of this subsection).

Third, having the way to calculate the left and right side of the equation, it is only necessary to estimate the parameters s_1 and s_2, in this case this parameters were estimated in the simulation stage for a gamma distribution, which are summarized in the table 3.3. Although it is possible to calibrate these parameters again, when applying the same ones used previously it is consistent with the construction of the simulations, which were randomized and all simulation is carried out trying to create representative samples of the distributions and therefore they should be values of s_1 and s_2 robust enough to apply in this context.

Once the MGF method is applied, the goodness of adjustment of the resulting distribution is estimated using the Kolmogrov-Smirnov test. This test is carried out against all data sets, one from a simulation based on the adjusted distribution and another from the actual data that are taken from the paths. In other words, we verify that a data set that follows the estimated distribution distributes sufficiently similar to the real data set that we want to adjust.The results of the adjustment obtain a positive result when applying the test in 93% of the cases, that is, 93% of the times we estimate the paths distributions it cannot be rejected that the estimated distribution distributes as the same as the real data . This 93% represents 686 paths of the original 735, which allows us to ensure the good performance of the methodology presented here. These results are summarized in the table below.

Acceptance percentage	Number of paths accepted
96,31%	683

Table 3.7: Results of the application of the method with the real data

To better understand the implication of approximating the sum of travel times using this methodology, the following graph shows the errors measured from the sample against the real travel time data.

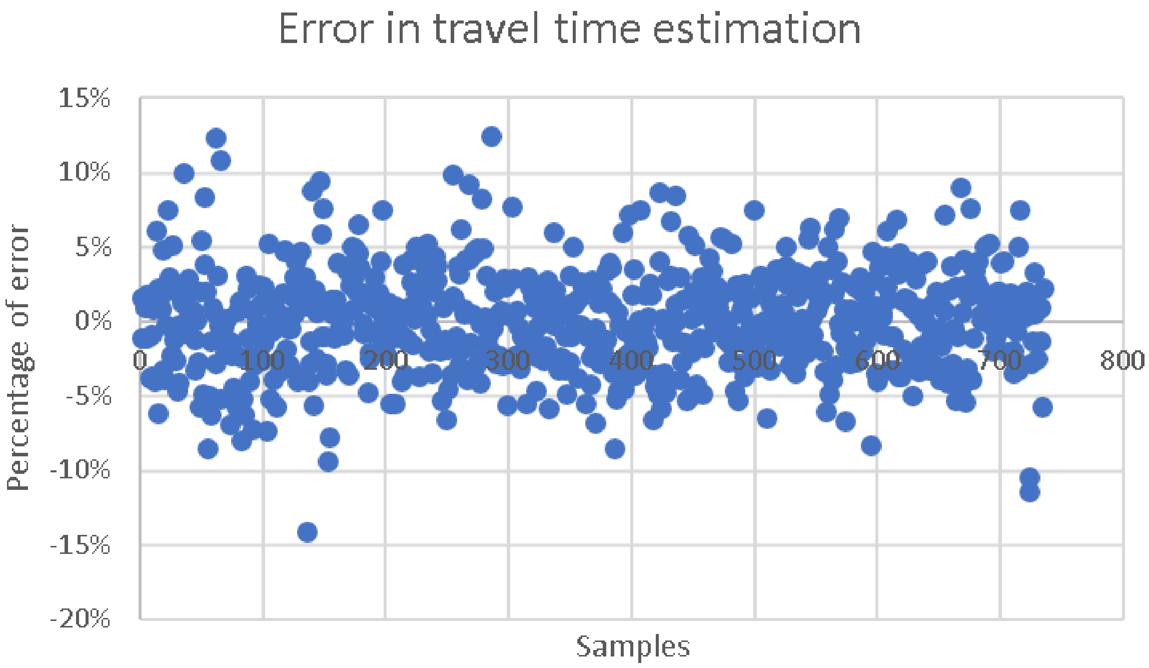

Figure 3.3: Error of estimation of travel time

Here it is observed that although some measurements are above 10% error, in general this value could be set as an error level, which is promising for the estimation of travel times. It should be noted that the travel times on the roads remain in a range of between 1,600 and 16,000 seconds, which will eventually lead to the real impact of this error.

4 Conclusion

Throughout this work it was observed that the gamma distribution works to be able to approximate the sum of gamma variables by the method of matching the MGF of both distributions. In this sense, it was developed how to implement this methodology for any distribution, subject to certain very general conditions. These conditions are:

1. The random variables correlated must have the same distribution
2. Know distribution function or density of the multivariate correlated random variables
3. Have a way to numerically estimate the distribution to which you want to approximate the sum of random variables, depending on its parameters

Having these 3 conditions, the methodology that was applied here was an example of how to implement it, where the distribution to which to approximate, as well as the validation of the adjustments made by the Kolmogrov-Smirnov test was developed without knowing at first what that would lead, but validating that following the form described in this work you can get a resounding result. In addition, each of the steps developed in the previous work were conceived, as processes, in a generic way, however, applied and tested specifically for gamma distribution. This processes were verified applying them in the approximation of the sum of log-normal correlated with a log-normal distribution, obtaining less than one 1% standard error when approximating the sum of log-normal.

Since the implementation was carried out allowing all the decisions of the steps to approximate the sum of the correlated variables to be free, that is, the results of the simulations themselves were the ones that guided the following steps, then it is understood that the results allow us to assume that before any set of correlated distributions this methodology can be implemented to obtain an approximation of the sum of random variables.

It is interesting to highlight the results obtained in Table 3.4, where it is observed that the performance of the methodology depends largely on the configuration that the underlying distributions of the path have, that is, it is not enough to observe what distributions are added but also how they are . This may imply that, for example, as we saw in general the sum of gamma correlated random variables can be approximated by a gamma distribution, however in the case that the summed gamma distributions have a parameter of form alpha = 1 and only 3 arcs are added , then perhaps there is a distribution that approximates better than a gamma, given the low result (at least with respect to the rest) of 73%.

That said, the question that remained was that although it worked in a controlled environment, of simulated variables that distributed exactly as we wanted, what would happen if it was applied in reality, where the big difference is that real data has a structure that resembles to a distribution, however, this distribution is not exactly as it would happen with simulated data, this could generate problems, given the change in the structure of the variables, however, when applying this form in section 3.5, a positive result was clearly obtained, which according to what is expected is worse in some configurations of the simulated variables, where 99% effectiveness was achieved against 93% obtained with real data, which is still very good.

It is important to highlight that although the results with real data are excellent, the paths that had the conditions that had been established previously were chosen, basically to perform a proof of concept of the above. This means that in order to test the structure developed with real data, those real structures were chosen that, statistically, follow those that were established previously, concluding that for a set of real data that behave in this way we could expect to have the reuse presented here.

It can be clearly said that the methodology of [6] is applicable, at least, for a gamma distribution, establishing the way in which it is widely implemented and therefore suggests that, in general, it should work as a way to solve the problem of the distribution of the sum of correlated random variables, at least applied in the form presented here, against other distributions, however the performance is subject to each case study due to the fundamental variations that remain in each part of the process (different multivariate distribution or quadrature used, for example)

4.1 Next Steps

To continue this work, it is interesting to take 3 different fronts:

1. Continue validating the methodology established here by taking another distribution or other data set with which to make this method more robust
2. Continue advancing in the application of this methodology, that is, start exploring applications of aversion to risk once the distribution is lifted, this implies evaluating situations such as change in travel time variability, increase in average travel time, establish metrics from aversion to risk, etc.
3. To extend this investigation in terms of the limitations presented here, these correspond to maintaining the same type of distribution throughout the graph and to know beforehand which distribution approximates the joint distribution of the arcs of the graph. A proposed line could be to use a distribution with more parameters that approximate the joint distributions and know which distribution to approximate this distribution, although this does not completely solve the problem, it does give more freedom

The above focuses basically to understand how to release this methodology from dependence to equality of correlated distributions or that the density function of multivariate distribution is not necessarily known. In addition to implementing this work in other contexts, to test it with other distributions and in other applications.

5 Bibliography

[1] I. Abramowitz, M. & Stegun. Handbook of mathematical functions: with formulas, graphs, and mathematical tables. *Courier Dover Publications*, 55, 1964.

[2] M-S Alouini, Ali Abdi, and Mostafa Kaveh. Sum of gamma variates and performance of wireless communication systems over nakagami-fading channels. *IEEE Transactions on Vehicular Technology*, 50(6):1471–1480, 2001.

[3] Philipp Arbenz, Paul Embrechts, and Giovanni Puccetti. The aep algorithm for the fast computation of the distribution of the sum of dependent random variables. *Bernoulli*, pages 562–591, 2011.

[4] Steven I-Jy Chien and Chandra Mouly Kuchipudi. Dynamic travel time prediction with real-time and historic data. *Journal of transportation engineering*, 129(6):608–616, 2003.

[5] Cristian E Cortés, Jaime Gibson, Antonio Gschwender, Marcela Munizaga, and Mauricio Zúñiga. Commercial bus speed diagnosis based on gps-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4):695–707, 2011.

[6] Daniel Espinoza M. Felipe Lagos G., Fernando Ordóñez P. Computing path travel time distributions. Master's thesis, Universidad de Chile, 2014.

[7] Marcello Galeotti. Computing the distribution of the sum of dependent random variables via overlapping hypercubes. *Decisions in Economics and Finance*, 38(2):231–255, 2015.

[8] Geoffrey Grimmett and Dominic Welsh. *Probability: an introduction.* Oxford University Press, 1989.

[9] Greg Kochanski. Monte carlo simulation. *URL www. ugrad. cs. ubc. ca/˜ cs405/montecarlo. pdf*, 2005.

[10] Dominik Kortschak and Hansjörg Albrecher. Asymptotic results for the sum of dependent non-identically distributed random variables. *Methodology and Computing in Applied Probability*, 11(3):279–306, 2009.

[11] Saralees Nadarajah. A review of results on sums of random variables. *Acta Applicandae Mathematicae*, 2008.

[12] AJ Richardson and MAP Taylor. Travel time variability on commuter journeys. *High*

Speed Ground Transportation Journal, 12(1), 1978.

[13] Ravi Seshadri and Karthik K Srinivasan. Algorithm for determining most reliable travel time path on network with normally distributed and correlated link travel times. *Transportation Research Record*, 2196(1):83–92, 2010.

[14] Karl Sigman. Acceptance-rejection method. *Columbia University*, 2007.