Development of statistics for cultural concepts measured by linguistic changes

Mirtha Haydee Pari Ruiz

Supervisor: Prof. Milan Stehlík

Instituto de Estadística
Universidad de Valparaíso

This dissertation is submitted for the degree of
Doctor en Estadística

Executive Summary

In the digital era, given so many data analysis techniques it is important to select an adequate one that ensures the quality of information. This thesis focuses on the distribution of quasi-distances of frequencies of linguistic objects selected from two historical corpuses of Nican Mopohua. The probability theory allows us to make statements in the presence of uncertainty, information allows us to quantify the amount of uncertainty in a given data. The theory of information has been applied to statistical and probabilistic problems with success in many research areas, also in linguistics and text comparisons, e.g. Bigi (2003) indicates in the context of linguistics that the Kullback-Leibler divergence is a measure of relative entropy that tells us how differ two probability distributions of linguistic object. Shannon, known as "the father of information theory," along with Warren Weaver, contributed to the culmination and settlement of the 1949 Mathematical Theory of Communication (now known as Information Theory). Divergences were widely studied e.g. by Kullback, Leibler and Rényi, among others. Divergences have multiple applications in signal and image processing, medical image analysis, texture classification, applications of natural language processing, etc. There exist divergence measures such as similarity functions or quasi-distances between two probability distributions that are: Kullbak-Leibler divergence (KL), Jensen-Shannon (JS), Skew divergence, Euclidean, cosine, L_1 and confusion, among others.
Lee (2001) explains that the KL is asymmetric measure of quasi-distance between two probability distributions, the Jensen-Shannon divergence is symmetric and considers the KL divergence between probabilities: 1) p and $\frac{p+q}{2}$ 2) q and $\frac{p+q}{2}$. The skew divergence is asymmetric quasi-distance. The Euclidean, the Cosine and the L_1 distance treat the distribution as vectors of relative frequencies and the Confusion probability estimates the substitutability of two given words.
Many investigations carried out on given linguistic corpus the analysis through frequency counting, it seems simple and it is a characteristic that allows us to calculate the probability, which conceptually is the analogue of the relative frequency.
Gutierrez and Cintas (2013) indicate that the histogram is an instrument for the understanding of probability density functions. On the other hand, if the model contemplates the theory of statistical information, it is more feasible to obtain point estimators to deal with Goodness-Of-Fit (GOF) tests with greater robustness obtained from divergence measures.

In this thesis we introduce information statistical quasi-distance to obtain an index of similarity or dissimilarity in the comparison of two probability distributions of the frequencies of evolved linguistic object, e.g. word in the corpus, measured in two distinct times.

Recognizing the great importance of language as the cornerstone of communication, it is difficult to know its origin according to the monogenetic hypothesis that all human languages derive from an ancestral language that must have appeared before the departure of homo sapiens from Africa. Being until today a mystery for many sciences, mainly the linguistic discipline tries to investigate this problem. Therefore, it is deduced that the origin is born in each nation by its naturalness, since the alphabet is being built with the passage of time and it is a tool of human knowledge. Ngugi wa Thiong'o, a recurring Kenyan writer nominated for the Nobel Prize for Literature, tells us that "Every language is a treasure of beauty, history, knowledge and possibilities". There are studies such as the Hilpert and Gries (2008) assessment of multiple changes over multiple periods of time addresses the need for a basic analytical toolbox that is specifically tailored to the interpretation of frequency changes in multistage diachronic corpora.
Amato et al. (2018) carried out a study of the dynamics of norm change in the cultural evolution of language. They investigated the process of norm change by looking at 2,541 linguistic norm shifts occurring over the last two centuries in English and Spanish. They identify different patterns of norm.

We use the methods of statistical information theory to determine a feasible model that detects linguistic variants for the interest of many other disciplines. But this research has a more epistemological background which serves other sciences and society since its theme of study incites the interest and reflection of identity, respect for the ancient cultures and the evolution of the study of the phenomenon of the change of the native language. Our research goes beyond a frequency report, adheres to means of quasi-distance and divergence methods and then makes a good adjustment to obtain a model of optimal distribution of the specifically defined variability of the native language.
According to the National Institute of Indigenous Languages (INAI) of Mexico the most spoken indigenous language is Nahuatl with 1,725,000 speakers (February, 2019)
The research carried out in this thesis is the analysis of linguistic data from the Nican Mopohua historical corpus (Nahuatl-Mexico Report) of two authors from different periods that of Lasso De La Vega (1649) and Rojas (1978) which constitute the stratified samples. We applied the principle of the aggregation (Stigler, 2016) in topological setup of Stehlík (2016). This turns out to be an advantage over the simplicity of symmetry distance based measures because the data confirmed empirically asymmetry. We applied several GOF tests to check the fit of uniform and gamma distribution to the empirical distribution of quasi-distances between both corpuses, based on carefully chosen linguistic benchmarks, e.g. keywords in order to identify the level of linguistic variation in time of the words in the Nahuatl language.

This defines a very ambitious research per se, since it is a great challenge to cover an ancestral language of the Mexica empire (Aztec empire) of the fourteenth century. Albeit the living version of language is still spoken in minority in Mexico we were attracted to select the classic Nahuatl since being one of the most documented and studied language in Americas.

It poses a series of variants since much of what seems to be only one language is due to the efforts of the friars who wanted to transmit, through the most established and cultured language, the Christian faith. Subsequently, the colonial authorities, with a more colloquial form, allowed the preparation of documents (testaments, territorial disputes, denunciations, etc.) in the Nahuatl language. It is also necessary to note that the spelling used by Lasso in the corpus presented five different diacritic: the macron $(\bar{a})$ for the long vowels; the accent $(\acute{a})$ for short vowels; the circumflex $(\hat{a})$ and the grave $(\grave{a})$ for the saltillo in different position and the cedilla (ç). The latter for the realization of data analysis were counted as a simple "a" and "c" for the report of frequencies in the determination of the alphabet (20 letters with digraphs) and for the counting of the classic Nahuatl keywords. It is also necessary to indicate that we set the geographic location of Mexico City since here the alphabet and the translation of the words were established by the expertise of the native teachers who helped us with several orthographic corrections and contributed to the design of the second experiment with the alphabet of 20 letters. There were also differences in letter and key word frequencies (lexeme + morpheme). The author of thesis visited the native group to obtain all these information.

Three experiments were carried out with the keywords, the first experiment consisting of study of 111 words, second experiment based on the 35 words that present lower frequency of occurrence and the third experiment analyses 74 words of the lowest and regular frequencies. In all cases the empirical probability of word occurrence was constructed individually for each corpus, and their quasi-distances with respect to seven quasi-distances were evaluated. Finally we checked hypothesized uniform and gamma distributions of quasi-distances by selected GOF tests.

We provide empirical justification for gamma distributed quasi-distances (divergences, geometric distances, etc.) between numerical characteristics of two corpuses of Nican Mopohua. This is not surprising, since gamma distribution is a very rich submodel of generalized gamma distribution family, which is approximating evolutions of stochastic differences. Such differences are generated in our subjective science by evolution of words in Nahuatl (e.g. semantic, cultural, morphological, etc.) (Klebaner et al., 1989). Here we use the fact that language semantic changes can be modeled as random walk with drift (Klebaner et al., 1989) (Hubey, 1999) Girosi and King has shown in 2007 that the Lee-Carter model is

equivalent to a special type of multivariate random walk with drift (RWD) model, in which the covariance matrix depends on the drift vector. These observations suggest that, since the RWD does not make any assumption about the structure of the covariance matrix, while the Lee-Carter approach does, the Lee-Carter estimator will be preferable to the RWD only when we have high confidence in its underlying assumptions. Such a model have a form

$$X_{t+1} = X_t + \theta + \varepsilon_{t+1}, \varepsilon_t \sim N(0, \sigma^2), \quad t \in [0, +\infty) \tag{1}$$

Let us generalize (1) by

$$X_{t+1} = X_t + g(X_t) + \eta_{t+1}, \tag{2}$$

where g is a positive function and $\{\eta_t\}$ is a square-integrable martingale difference sequence, the second conditional moments of which depend only on the present state of the process X_t. The state of the process X_t represents the given numerical characteristic of fixed keyword of corpus in time t where $t \in [0, +\infty)$

It is known that a large class of processes (Hubey, 1999) diverges with positive probability, and when properly normalized converges almost surely or converges in distribution to a normal or a lognormal distribution (here notice that one sided normal and lognormal distributions are special cases of a generalized gamma distribution (ggd)). Klebaner et al. (1989) has found a class of processes that when properly normalized converges in distribution to a ggd. Applications of this result to state dependent random walks and population size-dependent branching processes yield new results and reprove some of the known results. In such a setup usage of generalized gamma distribution is properly justified and Lee-Carter model could be a properly specified special case, if we are sure about its underlying assumptions. Notice, that beside this random walk justification, feasible estimation and testing procedures for ggd are developed, see e.g. Stehlík (2008) and references therein.

This thesis is organized as follows. Chapters 1 introduce both the preliminaries and selected more advanced properties of the divergence functions (Liese and Vajda, 1987) (Pardo, 2005). Also it gives introduction to some generalized measures of divergence and several notions of the most important distance measures (Hellinger and Rényi) and it gives brief summary on the Kerridge inaccuracy and the Divergence scores. Chapter 2 presents the study objectives that motivated and guided the research. We identify that written language Nahuatl is a very complex object due to its nature and structure, thus supporting the idea of topological aggregation and topological proximity of keywords. Some examples of words are presented, to illustrate the phenomenon of agglutination in the Nahuatl language. Then two experiments of occurrence of frequencies and the identification of the KL-divergence to

establish the Nahuatl alphabet follow. Application of GOF for the gamma distribution in first experiment is conducted.
The uniform distribution fit to empirical quasi-distances is conducted in Chapter 3. Several supporting simulations and results of two experiments with low and regular word frequency in corpus are provided.

Finally, in Appendix A we introduce the list of the 111 Nahuatl keywords which were studied, grouped according to their structure (lexeme and morpheme). In appendix B we place the packages and commands used in the statistical program R.

I dedicate this thesis to God, to the Virgin of Guadalupe, to my Father, to my Mother, to my sisters, to my fraternal friends and to my teacher friends who are part of my learning to learn to live wisely.

"Only those who try the absurd can achieve the impossible"

Albert Einstein

Acknowledgements

My thanks of open heart to my advisor, teacher and friend Prof. Milan Stehlik and his wife Silvia Stehlikova for their teaching, guidance, motivation, kindness, patience, advice and great human qualities worth imitating. Those with their wisdom helped me to complete and finish the thesis.

Thanks to all the professors of the Statistical Institute of the University of Valparaíso where each one of them contributed their experiences and knowledge and carved in me the love of knowledge and therefore of research. Also, I am grateful and from now on I will call my alma mater to the University of Valparaíso for their help and support, which made of my stay during these almost 3 and a half years the experience of a beautiful episode of my life of encounter with myself. Thank you

I thank the Pachamama of Chile for sheltering me and feeling at home always.

Abstract

Spanish Version

Poca o casi nada de investigación existe sobre cambios ortográficos o variación lingüística de una lengua nativa y los pocos estudios que existen se basan en un volumen de análisis de datos de muchos corpus de periodo de tiempo, detectando diferencias frecuenciales pero para determinada cantidad de palabras. Nuestra tesis para su estudio se basó en el corpus histórico Nican Mopohua (relato Náhuatl-México) de Lasso(1649) y Rojas(1978) y nos enfocamos en la comparación de la ocurrencia de frecuencias de palabras y adherido a esto se utiliza métodos de similaridad o divergencia en dos distribuciones de probabilidad para modelar el comportamiento de los datos. Específicamente, nos enfocamos en determinar la estadística formal para medir la casi distancia entre dos versiones del corpus para poder evaluar los cambios lingüísticos de la lengua Náhuatl en el tiempo. Además ajustamos un modelo estadístico que detectan variantes lingüísticas históricas y verificamos por pruebas de bondad de ajuste la distribución Gama y Uniforme como una distribución estadística adecuada de algunas cuasi distancias. Recopilamos nuestros datos respectivamente a los diseños de tres experimentos: 111 palabras (todas las palabras claves), 35 palabras (Palabras de menor frecuencia) y 74 palabras (palabras de menor y mediana frecuencia). En el primer diseño del experimento la distribución empírica de las cuasi distancias se ajusta bien a una distribución Gama, con la excepción de la medida de similitud de Coseno (distribución Beta) y Confusión (distribución de Cauchy). En el segundo experimento la distribución empírica de las cuasi distancias se ajusta a la distribución Uniforme y en el último experimento, la distribución empírica de las cuasi distancias también se ajusta a la distribución Uniforme, con la excepción de la medida del Coseno que tiene una distribución Beta como el mejor ajuste. La no presencia de normas léxicas y la naturaleza aglutinante de la lengua Náhuatl son grandes obstáculos. Nuestros resultados indican una justificación empírica de la distribución gama generalizada como modelo adecuado de cambios semánticos entre los dos corpuses del Nican Mopohua estos son presentados y pueden ser usados como ayuda para trabajos futuros.

English Version

Little or almost no research exists on orthographic changes or linguistic variation of a native language and the few studies that exist are based on a volume of data analysis of many corpuses of time period, detecting frequency differences for a certain number of words. Our thesis focuses on the historical corpus Nican Mopohua (narrative Náhuatl-México) of Lasso (1649) and Rojas (1978). We also focus on the comparison of the occurrence of word frequencies and use methods of similarity or divergence of two probability distributions to model the behavior of the data.

Specifically, we focused on determining the formal statistic for measuring of quasi-distance between two versions of the corpus in order to be able to asses the linguistic changes of the Nahuatl language in time. In addition we fit a statistical model to quasi-distances that detects historical linguistic variants and we verified by goodness-of-fit tests the Gamma and Uniform distribution as suitable statistical distribution of some quasi-distances. We collect our data respectively to the three designs of experiments: 111 words (all keywords), 35 words (words of lesser frequency) and 74 words (words of lower and medium frequency). In the first experiment, the empirical distribution of quasi-distances fits well to a gamma distribution with the exception of the cosine similarity measure (Beta distribution) and Confusion (Cauchy distribution). In the second experiment the empirical distribution of quasi-distances fits the uniform distribution and in the last experiment the empirical distribution of quasi-distances also fits the uniform distribution, with exception of the cosine measure that has a beta distribution as the best fit. The non-presence of lexical norms and the binding nature of the Nahuatl language are great obstacles. Our results indicate an empirical justification of generalized gamma distribution as suitable model of semantical changes between two corpuses of Nican Mopohua. These are presented and can be used as an aid for the future works.

Table of contents

List of figures **xv**

List of tables **xvii**

1 Preliminaries, f-divergences and related statistical distances **1**
1.1 On the modification of the Pearson statistic 1
1.2 Entropy measure 2
1.3 Bregman Scores 3
1.4 Phi-disparities 4
1.5 Gamma distribution 4
1.6 Generalized gamma distribution 7
1.7 f-divergences and related statistical quasi-distances 7
1.7.1 f-divergences 8
1.7.2 Kullback-Leibler divergence 12
1.7.3 ϕ- divergence 14
1.7.4 χ^2- divergence 16
1.7.5 Hellinger and Rényi distances 17
1.8 Kerridge inaccuracy 19
1.8.1 Properties of Inaccuracy 19
1.9 The Divergence score (DS) 20
1.9.1 Decomposition 21

2 Methodology and Results **23**
2.1 Main Objectives 23
2.2 Aggregation in topological spaces: topological closeness of linguistic objects 27
2.2.1 Examples on topological proximity of Náhuatl words 28
2.3 Nican Mopohua 31
2.4 Results of 111 keywords and gamma GOF tests 39

3 Close to uniform distribution **55**
3.1 Uniform distribution 55
3.1.1 Discrete uniform distribution 55
3.1.2 Continuous uniform distribution 56
3.1.3 Results of GOF test for uniform distribution for quasi-distances between two corpuses 56
3.2 Uniform designs: measures of Uniformity 90

4 Conclusions **93**
4.1 Formal statistics for the pseudo-distances between two linguistic corpuses . 93
4.2 Uniform distribution of pseudo-distance 94
4.3 Generalized gamma distribution of pseudo-distance 95

References **97**

Appendix A **101**

2 **105**

List of figures

2.1 Nearest neighbor of the word: **Quimo**lhuili (told him) 28
2.2 Nearest neighbor of the word: **Iyo**llo (your heart) 29
2.3 Nearest neighbor of the word: **Teo**tl (God) 29
2.4 Nearest neighbor of the word: **Oqui**cac (he hear) 29
2.5 Nearest neighbor of the word: **Cihuapill**i (maid) 30
2.6 Nearest neighbor of the word: **Tlacatl** (man) 30
2.7 Nearest neighbor of the word: **Xochitl** (flower) 30
2.8 gamma distribution: $D(p||q)$ 40
2.9 gamma distribution: $D(q||p)$ 42
2.10 gamma distribution: $JS(p,q)$ 44
2.11 gamma distribution: $JS(q,p)$ 45
2.12 gamma distribution: $S\alpha(p,q)$ 46
2.13 gamma distribution: $S\alpha(q,p)$ 47
2.14 gamma distribution: $euc(p,q)$ 48
2.15 beta distribution: $cos(p,q)$ 50
2.16 gamma distribution: $L_1(p,q)$ 51
2.17 cauchy distribution: $conf(p,q)$ 52

3.1 uniform distribution: $D(p||q)$ 58
3.2 Empirical and theoretical CDFs: $D(p||q)$ 59
3.3 uniform distribution: $D(q||p)$ 59
3.4 Empirical and theoretical CDFs: $D(q||p)$ 60
3.5 uniform distribution: $JS(p,q)$ 61
3.6 Empirical and theoretical CDFs: $JS(p,q)$ 62
3.7 uniform distribution: $JS(q,p)$ 62
3.8 Empirical and theoretical CDFs: $JS(q,p)$ 63
3.9 uniform distribution: $S_\alpha(p,q)$ 64
3.10 Empirical and theoretical CDFs: $S_\alpha(p,q)$ 65

3.11 uniform distribution: $S_\alpha(q,p)$ 65
3.12 Empirical and theoretical CDFs: $S_\alpha(q,p)$ 66
3.13 uniform distribution: $euc(p,q)$ 67
3.14 Empirical and theoretical CDFs: $euc(p,q)$ 68
3.15 uniform distribution: $cos(p,q)$ 68
3.16 Empirical and theoretical CDFs: $cos(p,q)$ 69
3.17 uniform distribution: $L_1(p,q)$ 70
3.18 Empirical and theoretical CDFs: $L_1(p,q)$ 71
3.19 uniform distribution: $conf(p,q)$ 72
3.20 uniform empirical and theoretical CDFs: $conf(p,q)$ 73
3.21 uniform distribution (74 words): $D(p||q)$ 75
3.22 Uniform empirical and theoretical CDFs (74 words): $D(p||q)$ 76
3.23 uniform distribution (74 words): $D(q||p)$ 76
3.24 Empirical and theoretical CDFs (74 words): $D(q||p)$ 77
3.25 uniform distribution (74 words): $JS(p,q)$ 78
3.26 Empirical and theoretical (74 words): $JS(p,q)$ 79
3.27 uniform distribution (74 words): $JS(q,p)$ 79
3.28 Empirical and theoretical CDFs (74 words): $JS(q,p)$ 80
3.29 uniform distribution (74 words): $S_\alpha(p,q)$ 81
3.30 Empirical and theoretical (74 words): $S_\alpha(p,q)$ 82
3.31 uniform distribution (74 words): $S_\alpha(q,p)$ 82
3.32 Empirical and theoretical (74 words): $S_\alpha(q,p)$ 83
3.33 uniform distribution (74 words): $euc(p,q)$ 84
3.34 Empirical and theoretical CDFs (74 words): $euc(p,q)$ 85
3.35 beta distribution (74 words): $cos(p,q)$ 85
3.36 Empirical and theoretical CDFs (74 words): $cos(p,q)$ 86
3.37 uniform distribution (74 words): $L_1(p,q)$ 87
3.38 Uniform empirical and theoretical CDFs (74 words): $L_1(p,q)$ 88
3.39 uniform distribution (74 words): $conf(p,q)$ 89
3.40 Empirical and theoretical (74 words): $conf(p,q)$ 90

List of tables

1.1 Contingency table for χ^2 test 2

2.1 Similarity functions for probability distributions 26
2.2 **A**: Corpus I (Lasso-1649) Vs Corpus II (Rojas-1978) 37
2.3 **B**: Corpus I (Lasso-1649) Vs Corpus II (Rojas-1978) 38

3.1 **Experiment 2**:Comparison of values functions for probability distributions 57
3.2 **Experiment 3**: Comparison of values functions for probability distributions 74

Chapter 1

Preliminaries, f-divergences and related statistical distances

1.1 On the modification of the Pearson statistic

One of primordial motivations for building up quasi-distance measures between two corpuses of historical text is the modification of the Pearson statistic. In 1900 Pearson created the χ^2 test that was the appropriate statistical test of independence for contingency tables with the aim of being able to verify the randomness of the discrepancies between the theoretical and the empirical distribution. Therefore, this test allows to find the asymptotic distribution for the known parameters, however the error of the estimation of unknown parameters, modification of the degrees of freedom and the theory of estimation for maximum likelihood, which allows to find the correct asymptotic distribution of the χ^2 was established by Fisher (1922) and confirmed by Yule (1922).
The statistical goodness-of-fit test originally suggested by Pearson (1904). The chi-squared test was:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \qquad where \quad E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

where O_i is the observed frequency, E_i is the expected frequency and N_i is the total frequency in corpus i (i takes the values 1 and 2 for the Lasso and Rojas corpora respectively) showing statistically significant difference at the 1% level.
In corpus Nican Mopohua, we use a 2x2 table to compare frequences of words between two corpora. The χ^2 test is a table with r rows and c columns. The number of degrees of freedom (d.f) is calculated by $(r-1)\cdot(c-1)$ this is equal to 1. Next, we show the 2x2 contingency

table in table 1.1

	Corpus one	Corpus 2	Total
Frequency of features	a	b	$a+b$
Frequency of features not ocurring	c	d	$c+d$
Total	$a+c$	$b+d$	$N=a+b+c+d$

Table 1.1 Contingency table for χ^2 test

So, we can calculate the χ^2 statistics as follows:

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Yates (1934) continuity correction, developed to improve the approximation of the continuous probability distribution (χ^2) to the discrete probability distribution of the observed frequency (multinomial). The Yate's corrrected chi-squared statistics (Y^2) is:

$$Y^2 = \frac{N(|ad-bc|-0.5N)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Other alternative is Fisher's exact test may be used for tables with small expected frequencies. It uses the observed frequencies themselves to find the probability (P) of obtaining any particular arrangement of frequencies a,b,c and d:

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}$$

1.2 Entropy measure

(Pardo, 2005) Let $\mathbf{X}$ be a random variable with probability distribution P_θ. From a historical perspective the first entropy measure was Shannon's entropy (1948), for continuous data

$$H(\mathbf{X}) \equiv H(P_\theta) \equiv H(\theta) = -\int_x f_\theta(x) log f_\theta(x) d\mu(x) = E_\theta[-log f_\theta(\mathbf{X})]$$

Where θ is the vector of unknown parameters. And for discrete data

$$H(\mathbf{X}) \equiv H(p_x) = -\sum_x p(x) log p(x)$$

1.3 Bregman Scores

(Kanamori and Sugiyama, 2014a) the Bregman score is an extension of the log-likelihood function (see Gneiting and Raftery (2007) for the details) (Bregman, 1967) (Pardo and Vajda, 1997). For functions f and g on $\mathbb{R}^d$, the Bregman score $S(f,g)$ is a class of real-valued functions that satisfy the inequality.

$$S(f,g) \geq S(f,f)$$

If the equality $S(f,g) = S(f,f)$ leads to $f = g$, $S(f,g)$ is called the strict Bregman Score. Where the $min_g S(f,g)$, has the uniquely optimal solution $g = f$.
For a function f defined on the Euclidean space $\mathbb{R}^d$, let $G(f)$ be a real-valued convex functional is called the lower potential. The functional derivative is $G'(x;f)$, which is defined as the function satisfying the equality

$$\lim_{\varepsilon \to \infty} \frac{G(f+\varepsilon h) - G(f)}{\varepsilon} = \int G'(x;f)h(x)m(dx),$$

for any function $h(x)$ with a regularity condition, where $m(.)$ is the measure. Then, the Bregman score $S(f,g)$ for functions f y g is

$$S(f,g) = -G(g) - \int G'(x,g)(f(x) - g(x))m(dx),$$

due to the convexity of $G(f)$, we have

$$G(f) - G(g) - \int G'(x,g)(f(x) - g(x))md(x) \geq 0,$$

that is equivalent to the inequality. Let $\mathscr{F}$ be a set of functions defined on $\mathbb{R}^d$ and if f is a probability density, the Bregman score is expressed as

$$S(f,g) = \int f(x)l(x,g)m(dx),$$

where $l(x,g)$ is given by

$$l(x,g) = -G'(x,g) - G(g) + \int G'(y,g)g(y)m(dy),$$

The function $l(x,g)$ is regarded as the loss of the forecast using $g \in \mathscr{F}$ for an outcome $x \in \mathbb{R}^d$.
Therefore, the Bregman score estimates the difference and the density ratio and the L_1-distance estimator is not significantly affected by extreme outliers.

1.4 Phi-disparities

We can consider quasi-distances as information Phi-disparities. This term appeared first in Lindsay et al. (1994), who has been developing statistical inference based on the ϕ-divergence, called ϕ-divergence statistics. This approach has several variants, e.g. replacing either one or both probability distributions by suitable estimators and require either bounded differentiability of ϕ or boundedness of ϕ, itself. Menéndez et al. (1998) introduced this term as an extension of the ϕ-divergence.

Definition The ϕ-disparity between the probability distributions P an Q is defined by

$$D_\phi(P,Q) = \int_\chi Q(x)\phi\left(\frac{P(x)}{Q(x)}\right) d\mu(x),$$

where the function $\phi : (0,\infty) \to [0,\infty)$ is assumed to be continuous, decreasing on $(0,1)$ and increasing on $(1,\infty)$, with $\phi(1) = 0$. The value $\phi(0) \in (0,\infty]$ is defined by the continuous extension.

Remark Note that the class of ϕ-disparities contains all ϕ-divergence of Csiszár (see Csiszár (1967), Liese and Vajda (1987))with $\phi : (0,\infty) \to (0,\infty)$ convex and equal to zero only at 1. Then, the assumed convexity and $\phi(1) = 0$ imply that

$$\frac{\phi(t) - \phi(1)}{t-1} = \frac{\phi(t)}{t-1}$$

Is nondecreasing in the domain $t > 0$. Therefore, $\phi(t)$ is increasing in the domain $t > 1$ unless $\phi(t) = 0$ in the interval $(1,t_1)$, and decreasing in the domain $0 < t < 1$ unless $\phi(t) = 0$ in the interval $(t_0,1)$.

In the next section we will study the gamma distribution that has its special importance in probability and statistics and that is widely used to model quantities that take positive values. Moreover, we have theoretical motivation given by limits of random walks, thus we conjecture that distribution of quasi-distances $d(p,q), d(q,p)$ can be well approximated by gamma distribution.

1.5 Gamma distribution

The gamma distribution is widely used in different disciplines to model continuous variables that are always positive. This distribution is useful for making a very good adjustments

to positive random variables possessing finite variance.
A random variable X follows the gamma distribution with positive parameters a, b and we denote it by $X \sim G(a,b)$.

(Villaseñor and González-Estrada, 2015) proposed a new property of the gamma distribution which can be used to obtain new parameter estimators, which have closed analytical expressions and root mean square error comparable with that of the maximum likelihood estimators.

Parameter estimation

Its probability density function is given by

$$f_x(x) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}, \quad x > 0, \quad a > 0, \quad b > 0 \tag{1.1}$$

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a $G(a,b)$ distribution. The maximum likelihood estimator (MLE) of a and b are the solution to the equations

$n\log b + n\psi(a) - \sum_{i=1}^{n} \log X_i = 0$
and

$$na - \sum_{i=1}^{n} X_i/b = 0 \tag{1.2}$$

where $\psi(a) = \Gamma'(a)/\Gamma(a)$. Since these equations canot be solved analytically, the Newton-Raphson algorithm can be used to find the MLE, which are denoted by $\hat{a}$ and $\hat{b}$.

Estimators obtained by the method of moments (MME) are $\tilde{a} = \bar{X}_n^2/S_n^2$ and $\tilde{b} = S_n^2/\bar{X}_n$, where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are the sample mean and the unbiased sample variance.
The notation $Y = X/b$ and $Z = \log X$ will be intensively used in the whole manuscript. Consider the following result.

Theorem 1. If $X \sim G(a,b)$, then for $k = 1,2,3$ we have
$E\{Y^k \log Y\} = (a+k-1)E\{Y^{k-1}\log Y\} + \frac{\Gamma(a+k-1)}{\Gamma(a)}$
Let $cov(X,Z)$ denote the covariance of X and Z. Notice that, from Theorem 1 with $k = 1, E\{Y\log Y\} - aE\{\log Y\} = 1$. That is, $cov(Y, \log Y) = 1$. Therefore, they have the following results, formulated in the Remark 1.

Remark 1. If $X \sim G(a,b)$ distribution, then $b = cov(X,Z)$

The cases $k = 2,3$ in Theorem 1 will be used later on for obtaining some properties of the Gamma distribution. Let $Z_i = \log X_i, i = 1, ..., n$, and $\bar{Z}_n = \sum_{i=1}^{n} Z_i/n$. Notice that by Remark

1 the sample covariance is an estimator for b, that is

$$\check{b}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Z_i - \bar{Z}_n) \tag{1.3}$$

On the other hand, by Eq (1.2), they have

$$\check{a}_n = \bar{X}_n/\check{b}_n \tag{1.4}$$

Thus, from (1.3) and (1.4), an estimator for the variance of the Gamma distribution, $\sigma^2 = ab^2$, turns out be

$$\check{\sigma}_n^2 = \bar{X}_n\check{b}_n \tag{1.5}$$

Notice that estimators $\check{a}_n$ and $\check{b}_n$ involve the sufficient statistics $(\bar{X}_n, \bar{Z}_n)$ for the parameters a and b. As a consequence of the delta method, it is well known that the method of moments estimators are consistent and asymptotically normally distributed.

A new test for Gamma distributions (Villaseñor and González-Estrada, 2015)

Let X be a random variable with continuous distribution function F. Let $\mathscr{F}_G$ denote the family of Gamma cumulative distribution functions with probability density function given in (1.1). They propose a test for the composite null hypotesis $H_0 : F \in \mathscr{F}_G$ versus the alternative hypothesis $H_1 : F \notin \mathscr{F}_G$, based on a random sample $X_1, X_2, ..., X_n$ of size n from F.

The decision rule of a goodness of fit test for H_0 can be based on the ratio of two estimators of the population variance, similarly to the well known Shapiro-Wilk statistic for testing normality. For instance, they propose the ratio $V_n = S_n^2/\check{\sigma}_n^2$ as a test statistic, where S_n^2 is the unbiased sample variance and $\check{\sigma}_n^2$ is given in (1.5). Under the null hypothesis, the values of V_n are expected to be close to one.

Let $S_{nXZ} = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_n)(Z_i - \bar{Z}_n)$, which is an equivalent notation for $\check{b}_n$. The following results provide the asymptotic null distribution of V_n.

Theorem 2. Let $\{X_n, Z_n\}$ be a sequence of independent random variables following a bivariate distribution with mean vector $\mu = (\mu_x, \mu_z)$ and covariance matrix $\sum = \begin{vmatrix} \sigma_X^2 & \sigma_{XZ} \\ \sigma_{XZ} & \sigma_Z^2 \end{vmatrix}$, if $V_n = \frac{S_n^2}{\bar{X}_n S_{nXZ}}$, then

$$\sqrt{n}(V_n - \frac{\sigma_X^2}{\mu_X \sigma_{XZ}}) \to N(0, \gamma^2)$$

Corollary 1

Under $H_0 : X \sim G(a,b)$, $\sqrt{n}(V_n - 1) \to N(0, \gamma^2)$, where $\gamma^2 = \frac{1}{a}\{1 + a(a\sigma_Z^2 - 1)\} > 0$

They propose a test based on the statistic

$$V_n^* = \sqrt{n\check{a}_n}(V_n - 1) \tag{1.6}$$

Which rejects H_0 if $V_n^* < k_{\alpha/2}$ or $V_n^* > k_{1-\alpha/2}$, where the critical values $k_{\alpha/2}$ and $k_{1-\alpha/2}$, corresponding to a test of size equal to $\alpha \in (0,1)$ satisfy the condition: $\alpha = sup_{\alpha>0} P(V_n^* < k_{\alpha/2}$ or $V_n^* > k_{1-\alpha/2}|H_0)$. Notice that we only maximize the type I error probability on the set of possible value of a since statistics V_n^* is scale invariant.

In the following section we introduce the flexible distribution in statistical literature that has as subfamilies the exponential, gamma and Weibull distribution is studied

1.6 Generalized gamma distribution

The generalized gamma distribution (ggd) was introduced by Stacy (1962) and also know as a power of a gamma distribution. (Stehlík, 2008) Consider that the observation y_i has probabilty density $f(y_i|\vartheta_i)$ and the sample density has the form $h(y|\vartheta) = f(y_1|\vartheta_1)f(y_2|\vartheta_2) \cdot \ldots \cdot f(y_n|\vartheta_n)$, where $\vartheta_1, \ldots, \vartheta_n$. In our case f is density from generalized gamma family of the form $(g.g.d.(\alpha, \beta, \sigma_i))$

$$f(y_i|\vartheta_i) = \frac{\alpha}{\sigma_i \Gamma(\frac{1+\beta}{\alpha})}(\frac{yi}{\sigma_i})^{\beta} exp(-(\frac{y_i}{\sigma_i})^{\alpha}) \qquad y > 0, \qquad \alpha, \beta, \sigma_i > 0$$

where $\vartheta_i = (\alpha, \beta, \sigma_i)$ and the parameters α and β are shape parameters and σ_i is a scale parameter of the distribution. In the especial case of $\beta = \alpha - 1$ the ggd is called a Weibull distribution and in case of $\alpha = 1$ we obtain the gamma distribution. This distribution does have the ability to mimic the attributes of other distributions such as the Weibull or lognormal, based on the values of the distribution's parameters. While the ggd is not often used to model life data by itself , its ability to behave like other more commonly used life distributions is sometimes used to determine which of those life distributions should be used to model a particular set of data.
This distribution has many applications also in reliability theory, engineering, physics and hydrology.

1.7 f-divergences and related statistical quasi-distances

The main objective of the theory of information is to quantify how much information is in the data. It is necessary to recognize the behavior of the function of their information through the

study of divergence measures that measure the changes that occur or that measure distance or measure how far away from each other are between two distributions of probability to be able to obtain a quality of estimation of the information.

This section talks about topics related to information theory, mathematical statistics and theory of probability as well as the convexity and continuity of f-divergences on spaces of probability measures.

Researchers like CSISZAR (1963) already introduced the χ^2-criterion of mathematical statistics, and the information theory of Shannon (1948) as the family of f- divergences. (Liese and Vajda, 1987)
Also, we will define the f- divergence of probability distributions. With the most important lemmas and corollaries and where f denotes a convex function from the class $k[0,\infty)$, $\mathbb{f}$ the corresponding convex function of two variables and $(\mathcal{X},\mathcal{A})$ a measurable space.It also establishes the extra values of: ${}^*\mathbb{R} = \mathbb{R}\bigcup\{\infty\}$ and $\mathbb{R}^* =^* \mathbb{R}\bigcup\{-\infty\}$.

Likewise, to analyze the Kullback-Leibler divergences and the χ^2- divergence parametrized by α, that presents a value of real number, $(\mathcal{X},\mathcal{A})$ an arbitrary measure space, μ and ν σ-finite measures from $S(\mathcal{A})$ dominated by $\rho \in S(\mathcal{A})$ and P, Q probability distributions from $P(\mathcal{A})$ dominated by $\mu \in S(\mathcal{A})$. Symbols p y q denote $d\mu/d\rho$ and $d\nu/d\rho$ or $dP/d\mu$ and $dQ/d\mu$, depending on the context. Also we consider, two statistical distances: The Hellinger distance and the Rényi distance.
Next, we consider the bases of the divergence measures with the f-divergence. Let $(\mathcal{X},\mathcal{A})$ is a measurable space. $S(\mathcal{A})$ is set of all $\sigma-$ finite measures over $\sigma-$ algebra $\mathcal{A}$ of subsets of space.

1.7.1 f-divergences

The symbol $\ll$ denotes the ususal domination and $\perp$ the singularity of measures. For P, $\mu \in S(\mathcal{A})$ related by $P \ll \mu$, the symbol $dP/d\mu$ denotes an arbitrary finite-valued version of the Randon-Nikodym derivative.

Definition

Let μ,ν, $\rho \in S(\mathcal{A})$ and let $\mu \ll \rho$ and $\nu \ll \rho$. Put $p = d\mu/d\rho$, $q = d\nu/d\rho$ and

$$f(\mu\|\nu) = \int \mathbb{f}(p,q)d\rho \tag{1.7}$$

whenever the integral exists. The expression $f(\mu\|\nu)$ is called an f- divergence of μ and ν

Lemma 1. *In the notation of (1.7) , it holds for all* $\mu,\nu \in S(\mathcal{A})$

$$f(\mu\|\nu) = \int_{\{q>0\}} f(p/q)d\nu + \mu(\{q=0\}).f(\infty)/\infty. \tag{1.8}$$

If in particular $\mu \ll \nu$*, then*

$$f(\mu\|\nu) = \int f\left(\frac{d\mu}{d\nu}\right) d\nu \tag{1.9}$$

This means that the left-hand expression exists in $\{-\infty\}\cup\mathbb{R}\cup\{\infty\}$ *iff so does the right-hand expression in wich case the two are equal.*

Further the proof of this lemma 1 can be found in (Liese and Vajda, 1987, page 12)

Liese and Vajda (1987) provide the following seven important properties of f- divergences and some with their respective corollaries:

Properties

Range of values theorem

(1) The f- divergence $f(P\|Q)$ of probability distribution P, $Q \in P(\mathcal{A})$ takes on values from the interval $[f(1), f(0)+f(\infty)/\infty]$. All values from this interval can be attained unless $\mathcal{A} = \{\phi, \mathcal{X}\}$.

(2) It holds $f(P\|Q) = f(1)$ or $f(P\|Q) = f(0)+f(\infty)/\infty$ if $P = Q$ or $P \perp Q$ respectively.

(3) If f is strictly convex at point 1 then $f(P\|Q) = f(1)$only if $P = Q$. If moreover $f(0)+f(\infty)/\infty < \infty$ then $f(0)+f(\infty)/\infty$ only if $P \perp Q$.

For the proof of this property see (Liese and Vajda, 1987, page 14)

Uniqueness and symmetry theorem Let $f, g \in k[0,\infty]$ the relation:

(1) $f(u) = g(u)+c(u-1)$ for all $u \in [0,\infty]$ and some $c \in \mathbb{R}$ implies

(2) $f(P||Q) = g(P||Q)$ for all P,$Q \in P(\mathcal{A})$

If $\mathcal{A} \neq \{0,x\}$ then the identity (2) implies (1).
The function defined by

(3)
$$f^*(u) = \begin{cases} f(\infty)/\infty & u = 0; \\ uf(\frac{1}{u}) & u \in (0,\infty) \end{cases}$$

Belongs to $k[0,\infty]$. It is strictly convex at $1/u$ iff f is strictly convex at $u \in (0,\infty)$ and

(4) $f^*(P||Q) = f(Q||P)$ for all $P,Q \in P(\mathcal{A})$
If $\mathcal{A} \neq \{\phi,x\}$ then the f-divergence is symmetric in the sense that (4) holds for $f^* = f$ iff (1) holds for $g = f^*$ defined by (3)

Proof. See proof in (Liese and Vajda, 1987, page 15)

Monotonicity theorem For every $P,Q \in P(\mathcal{A}) \Rightarrow (y,B)$ it holds

$$f(P||Q) \geq f(P*K||Q*k)$$

If K is sufficient for (P,Q) then the sign of equality takes place. Conversely, if f is strictly convex at every $v \in (0,\infty)$ and

$$f(P||Q) = f(P*K||Q*k) < \infty$$

then k is sufficient for (P,Q).

Proof. See proof in (Liese and Vajda, 1987, page 17)

Approximation theorem Let $\mathcal{A}_\lambda \subset \mathcal{A}$ be a non-decreasing sub σ- algebras for λ from a directed set Λ, let

$$\mathbb{A} = \bigsqcup_{\lambda\in\Lambda} \mathcal{A}_\lambda \quad \textit{and} \quad \sigma(\mathbb{A}) = \mathcal{A},$$

and let P_λ, Q_λ be restrictions of $P,Q \in P(\mathcal{A})$ on $\mathcal{A}\lambda$. Then $f(P_\lambda||Q_\lambda)$ tends on λ non-decreasingly to $f(P||Q)$.
Proof. See proof in (Liese and Vajda, 1987, page 20)

Corollary 1. *Let $\mathbb{A}$ be a subalgebra of $\mathcal{A}$ generating $\mathcal{A}$, $\Lambda(\mathbb{A})$ the set of all partitions of $\mathcal{X}$ into a finite number of sets $\mathcal{A} \in \mathbb{A}$, $\mathcal{A}_\lambda = \sigma(\lambda) \subset \mathbb{A}$ for every $\lambda \in \Lambda(\mathbb{A})$, and let P_λ , Q_λ be restrictions of P, $Q \in P(\mathcal{A})$ on $\mathcal{A}_\lambda$.*

Isomorphy theorem Let $(x,\mathcal{A})$, $(y,\mathcal{B})$ be measurable spaces. If $(P_1,P_2) \in P(\mathcal{A})xP(\mathcal{A})$ and $(Q_1,Q_2) \in P(\mathcal{B})xP(\mathcal{B})$ are isomorphic pairs then

$$f(P_1||P_2) = f(Q_1||Q_2)$$

Lower semicontinuity theorem Let the weak topology τ_ϕ induced by ϕ be complete. Then the f-divergence is lower semicontinuous on the topological space $(P(\mathcal{A})xP(\mathcal{A}), \tau_\phi x\tau_\phi)$. *Proof.* See proof in (Liese and Vajda, 1987, page 27)

Convexity theorem For every $P \in P(\tau)$ and all kernels $k \equiv (P_\theta, \theta \in \theta)$, $\tilde{k} \equiv (Q_\theta, \theta \in \theta)$ of the type $(\theta,\tau) \Rightarrow (x,\mathcal{A})$ it holds

$$f(P*k||P*\tilde{k}) \leq \int f(P'_\theta||Q'_\theta)P(d_\theta)$$

where $k' \equiv (P'_\theta, \theta \in \theta)$ are restrictions of the kernels k, $\tilde{k}$ on the above introduced $\mathcal{A}'$. The equality takes place if $\{\phi,\Theta\} \otimes \mathcal{A}'$ is sufficient for (PxK', Pxk'). If f is strictly convex at all $u \in (0,\infty)$ then the last statement holds with "if" repaced by "iff" provided $f(P*K|P*\tilde{k}) < \infty$

Proof. See proof in (Liese and Vajda, 1987, page 31)

Corollary 2. *The f-divergence is a convex function of probability distributions in the sense that for every measurable space $(\mathcal{X},\mathcal{A})$ and all pairs of distributions$(P_i,Q_i) \in P(\mathcal{A})xP(\mathcal{A})$, $i = 1,2$, and for all $v \in [0,1]$ it holds.*

$$f(vP_1 + (1-v)P_2\|vQ_1 + (1-v)Q_2) \leqslant vf(P_1\|Q_1) + (1-v)f(P_2\|Q_2)$$

If P_i, $Q_i \ll \mu \in S(\mathcal{A})$ and p_i, q_i are the densities then the equality takes place if.

$$p_1q_2 = q_1p_2$$

$\mu - a.s$ In the case that f is strictly convex at all points from $(0,\infty)$ the last statement holds with "if" replaced by "iff" provided the values are finite.

Next we will analyze one of the most important divergence measures between two probability distributions, the Kullback-Leibler divergences which is basic concept of information theory.

1.7.2 Kullback-Leibler divergence

It is one of the most important families of divergences. In probability theory it is considered an indicator of the similarity between two distribution functions. (Kullback and Leibler, 1951) (Kullback, 1997). In information theory, it is known as information divergence or relative entropy $D(P||Q)$ which is a measure of the "distance" between two probability mass functions P and Q. It arises as an expected logarithm of the likelihood ratio. It is also considered as the measure of the inefficiency of assuming that the distribution is q when the true distribution is p. And It is defined for the continuous case as

$$\begin{aligned} D(P\|Q) &= \int_x p(x) \quad \log\left(\frac{p(x)}{q(x)}\right) dx \\ &= E_p\left[\log\left(\frac{p}{q}\right)\right] \end{aligned} \tag{1.10}$$

And for the discrete case it is

$$\begin{aligned} D(P\|Q) &= \sum_{x\in X} P(x)\log\left(\frac{p(x)}{q(x)}\right) \\ &= E_p\left[\log\left(\frac{p}{q}\right)\right] \end{aligned} \tag{1.11}$$

Where we use the convention that $0\log\frac{0}{0} = 0$ and the convention (based on continuity arguments) that $0\log\frac{0}{q} = 0$ and $p\log\frac{p}{0}0 = \infty$. Thus, if there is any symbol $x \in X$such that $p(x) > 0$ and $q(x) = 0$, then $D(P||Q) = \infty$. (Cover and Thomas, 2012)
The investigation is based on the discrete case already that the nature of the study variables so requires. Although relative entropy is not a true metric, it has some of the properties of a metric. In particular, it is always nonnegative and is zero if and only if $P = Q$. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality.Therefore, this measure quantifies the proximity of two probability distributions and we call such operator the quasi-distance in the thesis.

Next we will see the properties.

Properties :

According to Cover and Thomas (2012) the properties are:

(1) $D(P||Q) \geq 0$ with equality if and only if $p(x) = q(x)$, for all $x \in X$.

(2) $D(P||Q)$ is convex in the pair (P,Q)

(3) The KL- divergence is asymetric , i.e. $D(P||Q) \neq D(Q||P)$

The KL-divergence in the information measures I(1:2) and I(2:1)are as mean information for discrimination.
We consider the function $I_\alpha : [0,\infty) \mapsto^* \mathbb{R}$ (Liese and Vajda, 1987) defined by

$$I_\alpha(u) = \begin{cases} -lnu + u - 1 & \alpha = 0; \\ \frac{u^\alpha - \alpha u + \alpha - 1}{\alpha(\alpha-1)} & \text{if } \alpha \neq 0, \alpha \neq 1; \\ ulnu - u + 1 & \alpha = 1. \end{cases} \tag{1.12}$$

Lemma 2. *(Liese and Vajda, 1987) The function I_α is non-negative,*

$$I_\alpha(1) = 0 \quad and \quad I_\alpha(0) + I_\alpha(\infty)/\infty = \begin{cases} \frac{1}{\alpha(1-\alpha)} & if \quad 0 < \alpha < 1; \\ \alpha & otherwise, \end{cases}$$

Where it belongs to the class of convex functions $k[0,\infty)$.

Lemma 3. *(Liese and Vajda, 1987) The function $I^*_\alpha \in k[0,\alpha)$ conjugated to I_α satisfies the relation $I^*_\alpha = I_{1-\alpha}$*

Therefore, the f- divergence defined by 1.7 for $f(u) = I_\alpha(u)$ is called the I_α- divergence and denoted by $I_\alpha(\mu\|\nu)$ for all $\mu, \nu \in S(\mathcal{A})$
So, the non-negativity of I_α implies that $I_\alpha(\mu\|\nu)$ exists for all μ, $\nu \in S(\mathcal{A})$. Further, it follows from lema (1.8) that

$$I_1(\mu\|\nu) = \int_{pq>0} \left(p \ln \frac{p}{q} - p + q \right) d\rho + \infty.\mu(\{q=0\})$$

,

$$I_0(\mu\|\nu) = I_1(\nu\|\mu),$$

$$I_\alpha(\mu\|\nu) = \frac{1}{\alpha(\alpha-1)} \int_{\{pq>0\}} (p^\alpha q^{1-\alpha} - \alpha p - (1-\alpha) q) d\rho$$

$$+ \mathbb{1}_{(0,1)}(\alpha) \left(\frac{\nu(\{p=0\})}{\alpha} + \frac{\mu(\{q=0\})}{1-\alpha} \right)$$

$$+ \infty \mathbb{1}_{(-\infty,0)}(\alpha) \nu(\{p=0\}) + \infty \mathbb{1}_{(1,\infty)}(\alpha) \mu(\{q=0\})$$

If P, $Q \in P(\mathcal{A})$, i.e. if P, Q are probability distributions, then the formulas above simplify, e.g.

$$I_1(P\|Q) = \int_{\{q>0\}} ln\frac{p}{q}dP + \infty.P(\{q=0\})$$

And by the lemma(3) and the general symmetry theorem.

$$I_\alpha(Q\|P) = I_{1-\alpha}(P\|Q) \quad for \quad all \quad P,Q \in P(\mathcal{A}) \tag{1.13}$$

Proposition 1. *(Liese and Vajda, 1987) Let P, $Q \in P(\mathcal{A})$. If $\alpha \in (0,1)$ then*

$$0 \leq I_\alpha(P\|Q) \leq \frac{1}{\alpha(1-\alpha)}$$

Where the left equality holds iff $P = Q$ and the right equality holds iff $P \perp Q$ If $\alpha \notin (0,1)$ then

$$0 \leq I_\alpha(P\|Q) \leq \infty$$

Where the left equality holds iff $P = Q$ and the right one holds if $P \not<< Q (Q \not<< P)$ for $\alpha > 1 (\alpha < 0)$

1.7.3 ϕ- divergence

(Pardo, 2005) (Morales et al., 1996) The ϕ- measure between the probability P and Q is defined in the continuous case by

$$\begin{aligned} D_\phi(P,Q) &= \int_X q(x)\phi\left(\frac{p(x)}{q(x)}\right)du(x) \\ &= E_Q\left[\phi\left(\frac{p}{q}\right)\right], \quad \phi \in \Phi^* \end{aligned}$$

and in the discrete case by

$$D_\phi(P,Q) = \sum_x q(x)\phi\left(\frac{p(x)}{q(x)}\right)$$

where Φ^* is the class of all convex functions $\phi(x)$, $x \geq 0$, such that a $x = 1$, $\phi(1) = 0$ at $x = 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{x\to\infty}\phi(x)/x$

Remark 2.3 (Pardo, 2005)
Let $\phi \in \Phi^*$ be differentiable at $x = 1$, then the function

$$\psi(x) \equiv \phi(x) - \phi'(1)(x-1)$$

also belongs to Φ^* and has the additional property that $\psi'(1) = 0$. This property, together with the convexity, implies that $\psi(x) \geq 0$, for any $x \geq 0$. Further consider the set Φ^* to be equivalent to the set

$$\Phi \equiv \Phi^* \cap \{\phi : \phi'(1) = 0\}$$

The role of divergences obtained for $\phi(x) = -\log x$ and $\phi(x) = x\log x$ (Pardo, 2005) and the Kullback- Leibler divergence measure also is obtained for $\psi(x) = x\log x - x + 1$ or $\phi(x) = x\log x$ is well known in information theory (see also Stehlík (2003)).
(Pardo, 2005) The basic properties of ϕ-divergence are:

1 Let P_{θ_1} and P_{θ_2} be two probability distributions and let $\phi \in \Phi^*$ be differentiable at $t = 1$. Then:

$$0 \leqslant D_\phi(\theta_1, \theta_2) \leqslant \phi(0) + \lim_{r\to\infty} \frac{\phi(r)}{r}$$

where

$$D_\phi(\theta_1, \theta_2) = 0. \quad if \quad P_{\theta_1} = P_{\theta_2} \tag{1.14}$$

and

$$D_\phi(\theta_1, \theta_2) = \phi(0) + \lim_{r\to\infty} \frac{\phi(r)}{r} \quad if \quad S_1 \bigcap S_2 = \emptyset \tag{1.15}$$

If ϕ is also strictly convex at $t = 1$, then (1.13) holds if and only if $P_{\theta_1} = P_{\theta_2}$. If moreover:

$$\phi(0) + \lim_{r\to\infty} \frac{\phi(r)}{r} < \infty$$

then (1.14) holds if and only if $S_1 \bigcap S_2 = \emptyset$, where $S_i, i = 1,2$ is the support of the probability distribution P_{θ_i}, $i = 1,2$

Let $X_1, ..., X_n$ be a sample from P_θ, $\theta \in \Theta$. For μ being the Lebesgue measure or a counting measure, let $f_\theta(x) = \frac{dP_\theta}{d\mu}(x)$ where $x = (x_1, ..., x_n)$. Suppose that T is a measurable transformation from $(X^n, \beta x^n)$ onto a measurable space (Y, β_y). We denote

$$Q_{\theta_i}(A) = P_{\theta_i}(T^{-1}(A)), \quad i = 1,2, \tag{1.16}$$

With $A \in \beta_y$ and

$$g_{\theta_i}(t) = \frac{dQ_{\theta_i}}{d\mu}(t), \quad f_{\theta_i}(x/t) = \frac{dP_{\theta_i}}{dQ_{\theta_i}}, \quad i = 1,2; \tag{1.17}$$

by t we are denoting the values of T. In this context we have the following property.

2 Let $\phi \in \Phi^*$ and $Q_{\theta_i}, P_{\theta_i}$, $i = 1,2$, be two probability measures defined in (1.13) and (1.14). Then we have

$$D_\phi(Q_{\theta_1}, Q_{\theta_2}) \leqslant D_\phi(P_{\theta_1}, P_{\theta_2})$$

The equality holds if T is sufficient for the probability distributions P_{θ_1} and P_{θ_2}

In the following part we will talk about the χ^2 divergences wich is a vital concept of mathematical statistics and the total variation which has origins in measure theory.
We'll also see that the $\chi^2(P\|Q)$ go back to Pearson (1900).

1.7.4 χ^2- divergence

We restict ourselves to $\alpha > 0$ only. We consider the function $\chi^\alpha : [0,\infty) \mapsto^* \mathbb{R}$ defined by (see (Liese and Vajda, 1987))

$$\chi^\alpha(u) = \begin{cases} |1-u^\alpha|^{1/\alpha} & \text{if} \alpha \leqslant 1 \\ |1-u|^\alpha & \text{if } \alpha > 1. \end{cases} \tag{1.18}$$

Proposition 2. *(Liese and Vajda, 1987) It holds $\chi^\alpha \in k[0,\infty)$, χ^α is strictly convex at every point $u \in (0,\infty)$, unless $\alpha = 1$ when it is strictly convex at $u = 1$ only, and*

$$\chi^\alpha(0) = 1, \chi^\alpha(1) = 0, \chi^\alpha(\infty)/\infty = \begin{cases} 1 & if \quad \alpha \leq 1 \\ \infty & if \quad \alpha > 1 \end{cases}$$

The f- divergence defined by (1.7) for $f(u) = \chi^\alpha(u)$ is called a χ^α- divergence and denoted by $\chi^\alpha(P\|Q)$.
By lemma (1) and (1.18), it holds in the notation of (1.7) for all $P, Q \in P(\mathcal{A})$

$$\chi^\alpha(P\|Q) = \begin{cases} \int |p^\alpha - q^\alpha|^{1/\alpha} d\rho & \text{if} \alpha \leqslant 1 \\ \int_{\{q>0\}} |1-\frac{p}{q}|^\alpha dQ + P(\{q=0\})\infty & \text{if } \alpha > 1. \end{cases} \tag{1.19}$$

it is necessary to notice that the χ^1- divergence plays a similar role in the family of χ^α - divergences as the I_1- divergence does in the family of I_α- divergences as the I_1-divergence did in the family of I_α- divergences. So by (1.18) and corollary (1), it satisfies the relation.

$$\chi^1(P\|Q) = 2sup\{P(A) - Q(A) : A \in \mathcal{A}\} = V(P\|Q),$$

It is the total variation of signed measure $P-Q$ on $\mathcal{X}$.
It is an interesting situation that the families of I_α-, H_α- and χ^α- divergences intersect for certain special α. In addition to the relations

$$R_\alpha(P\|Q) = I_\alpha(P\|Q) \quad for \quad \alpha = 0 \quad and \quad \alpha = 1$$

Imposed directly in the definition of the Rényi's distance, the definitions given above yield the following relations valid for all $P, Q \in P(\mathcal{A})$:

$$\chi^{\frac{1}{2}}(P\|Q) = \frac{1}{2} I_{\frac{1}{2}}(P\|Q) = 2(1 - H_{\frac{1}{2}}(P\|Q))$$
$$\chi^2(P\|Q) = 2I_2(P\|Q) = H_2(P\|Q) - 1$$

We can see that the class of χ^α- divergences intersects with the class of I_α- divergences at divergences of extremal theoretical and practical importance.
Also with the proposition (2), one can transfer the basic properties of an f- divergence, where the range of values of the χ^α- divergence is the interval

$$[\chi^\alpha(1), \chi^\alpha(0) + \chi^\alpha(\infty)/\infty] = \begin{cases} [0,2] & \text{if } \alpha \leq 1 \\ [0,\infty] & \text{if } \alpha > 1. \end{cases} \tag{1.20}$$

The particular case is obtained for $\phi(x) = x \ln x$. Other well-known examples, the χ^2- divergene and modified χ^2- divergence (compare Pearson, 1900).

$$\chi^2(P,Q) = \sum_x \frac{(p(x) - q(x))^2}{q(x)} \quad and \quad \tilde{\chi}^2(P,Q) = \chi^2(Q,P) = \sum_x \frac{(p(x) - q(x))^2}{p(x)}$$

(Morales et al., 1996) are obtained for $\phi(x) = (x-1)^2$ and $\phi(x) = (x-1)^2/x$

Now, we continue with the distances Hellinger and Rényi.

1.7.5 Hellinger and Rényi distances

The Hellinger's integral of order α is the function defined for every P, $Q \in P(\mathcal{A})$ (Liese and Vajda, 1987)

$$H_\alpha(P\|Q) = \begin{cases} P(\{q > 0\}) & \alpha = 1; \\ 1 + \alpha(\alpha - 1) I_\alpha(P\|Q) & \text{if } \alpha \neq 0, \alpha \neq 1; \\ Q(\{p > 0\}) & \alpha = 0 \end{cases} \tag{1.21}$$

(Liese and Vajda, 1987) The next function is the Rényi's distance of order α, where $\ln 0 = -\infty$.

$$R_\alpha(P\|Q) = \begin{cases} I_0(P\|Q) & \alpha = 0; \\ \frac{1}{\alpha(\alpha-1)} ln H_\alpha(P\|Q) & \text{if } \alpha \neq 0, \alpha \neq 1 \\ I_1(P\|Q) & \alpha = 1 \end{cases}$$

the proposition(1) implies that the range of values is always nonnegative, that $0 \leq H_\alpha(P\|Q) \leq 1$ if $\alpha \in (0,1)$ and $H_\alpha(P\|Q) \geq 1$ for $\alpha \notin (0,1)$.
that is to say that in the extended form when $\alpha \neq 1, 0$ we have

$$R_\alpha(P||Q) = \frac{1}{\alpha(\alpha-1)} \log \int_x p(x)^\alpha q(x)^{1-r} d\mu(x)$$

$$\frac{1}{\alpha(\alpha-1)} \log E_p \left[\left(\frac{p(x)}{q(x)} \right)^{\alpha-1} \right]$$

Analogically one can easily obtain the condition to obtain equalities from these inequalities. Further, analogically, it holds
$0 \leq R_\alpha(P\|Q) \leq \infty$ and $R_\alpha(P\|Q) = \infty$ for $\alpha \in (0,1)$ iff $P \perp Q$.
Also, for the non-negativity of I_α one obtains similar formulas for the Hellinger integrals under consideration (Liese and Vajda, 1987). It holds

$$H_\alpha(P\|Q) = \int_{\{pq>0\}} (p^\alpha q^{1-\alpha} d\mu + \infty \mathbb{1}_{(-\infty,0)}(\alpha) Q(\{p=0\}) + \infty \mathbb{1}_{(1,\infty)}(\alpha) P(\{q=0\})$$

Here follows the symmetry theorem for I_α it holds for every $P, Q \in P(\mathcal{A})$ (Liese and Vajda, 1987)

$$H_\alpha(Q\|P) = H_{1-\alpha}(P\|Q), \quad R_\alpha(Q\|P) = R_{1-\alpha}(P\|Q)$$

Therefore, the solution to some classical problems of statistical inference, basically problems of estimation, can be found on the basis of measures of the divergence, with applications to statistical analysis of categorical data based in functionals of Information Theory. Thus it can be called Statistical Information Theory. The Minimum divergence estimators or minimum distance estimators (see William (1981)) have been used successfully in models for continuous and discrete data. Being so these statistics are very good alternatives to the classical Pearson—type statistic, in discrete models.

1.8 Kerridge inaccuracy

(Kerridge, 1961) The inaccuracy of this statement can be measured by $-\sum f_1(x)\log f_2(x)$
Let be the measure of inaccuracy

$$I = I(f_1, f_2)$$

Then we asume:

a) The functin I is continuous in the f_1 and f_2.

b) When N equally likely alternatives are stated , the inaccuracy is a monotonic increasing function of N.

c) If a statement is broken down into a number of subsidiary statements, the inaccuracy of the original statement is a weighted sum of the inaccuracies of the subsidiary statements.

d) The inaccuracy of a statement is unchanged if two alternatives about wich the same assertion is made are combined.

So, to carry out the derivation in almost exactly the way used by Shannon helps to show the very close parallel between entropy and inaccuracy. Therefore

$$I = -K\sum f_1(x)\log f_2(x)$$

or,

$$I(f_1, f_2) = K(f_1, f_2) = -\int_x f_1(x)\log f_2(x)dx$$

where $K = 1$. So, $-\sum f_1(x)\log f_2(x) = -\sum f_1(x)\log f_1(x) - \sum f_1 \log(f_2(x)/f_1(x))$
Or,

$$-\int_x f_1(x)\log f_2(x)dx = -\int_x f_1(x)\log f_1(x)dx - \int_x f_1(x)\log\frac{f_2(x)}{f_1(x)}dx$$

Where the first term represent the inaccuracy of uncertainty, and the second term the inaccuracy of error. This is hold by the fact that the error term vanishes when $f_1(x) = f_2(x)$

1.8.1 Properties of Inaccuracy

Kerridge (1961) presents the following properties:

a) The quantity I is zero if, and only if, $f_1(x) = f_2(x) = 1$ for one value, and conseuently $f_1(x) = f_2(x) = 0$ for all other value.The answer zero means a correct statement.

b) There is an infinite value of I if $f_2(x) = 0$, $f_1(x) \neq 0$ for any value. The result implies that truth is regarded as infinitely valuable.

c) The value of I is a minimum for fixed $f_1(x)$ when $f_2(x) = f_1(x)$, so that the error term is zero.It then reduces to the ordinary communication theory uncertainty.

d) If both $f_2(x)$ and $f_1(x)$ vary,the point $f_1(x) = f_2(x) = n^{-1}$ is a minimax point.

e) If two sets of alternatives are aserted to have probabilities which are independent, the inaccuracy of the joint assertion is the sum of the separate innacuracies.

Also, we can affirm that the uncertainty of a joint event is not greater than the sum of separate uncertainties, but this is not true of the inaccuracy. If the interaction between the events is wrongly stated, the joint event may have greater inaccuracy.
The following section explains Weijs et al. (2010) a score that can be used for evaluating probabilistic forecasts of multicategory events.

1.9 The Divergence score (DS)

Replacing the cuadratic distance from the Brier score (*BS*) with the Kullback-Leibler divergence Weijs et al. (2010) propose measure the forecast verification. This score is called Divergence score (*DS*) that is a function as a scoring rule. So, the forescast distribution from the observation distribution over the possible events i is

$$DS = D_{KL}(\mathbf{o}_t \parallel \mathbf{f}_t) = \sum_{i=1}^{n} [\mathbf{o}_t]_i \log(\frac{[\mathbf{o}_t]_i}{[\mathbf{f}_t]_i})$$

Also, for pairs measures of forecast observation the average divergence of the forecast distribution is:

$$DS = \frac{1}{N} \sum_{t=1}^{N} D_{KL}(\mathbf{o}_t \parallel \mathbf{f}_t)$$

where, $\mathbf{o}_t$ is the observation and $\mathbf{f}_t$ is a forecast.
This measure can be interpreted as the information gain when is the prior forecast distribution. When this value is zero it tells us that the forecast already contained all the information.
Next we show the similar descomposed of the measure Brier Score (*BS*) into three components.Murphy (1973)

1.9.1 Decomposition

(Weijs et al., 2010) The decomposition into three components provide us a diagnostic information about the quality of forecast
The relation between the components and the total score (DS) is:
$DS = REL - RES + UNC$

that is to say,

$$DS = \frac{1}{N}\sum_{k=1}^{K} n_k D_{KL}(\bar{\mathbf{o}}_k \parallel \mathbf{f}_k) - \frac{1}{N}\sum_{k=1}^{K} n_k D_{KL}(\bar{\mathbf{o}}_k \parallel \bar{\mathbf{o}}) + \sum_{i=1}^{n} |\bar{\mathbf{o}}|_i \log |\bar{\mathbf{o}}|_i$$

where N is the total number of forecast issued, K is the number of unique forecast issued, $\bar{\mathbf{o}} = \sum_{t=1}^{N} \mathbf{o}/N$, n_k is the number of forecast with the same probability category, $\bar{\mathbf{o}}_k$ is the observed frequency and $\mathbf{f}_k$ s the forecasts probability.

The first component, reliability (REL)measures the bias in the probability estimates. Then the resolution (RES) tell us how much uncertainty is explained by the forecast and finally the last component the uncertainty (UNC) the inherent uncertanty in the process and is measured by the entropy. This no depend on the forecast. And a uncertainty is maximum if the probability of ocurrence is 0.5 and 0 if the probability is 0 or 1. A perfect forecast has a resolution that is equal to the uncertainty and a perfect reliability.

Chapter 2

Methodology and Results

The present thesis aims to study a formal statistic for comparison between two corpuses of historical linguistic text. This is the first step in order to analyze the changes of the historical linguistic corpus by developing a statistical model for information quasi-distances which can detect spelling variations in the historical linguistic corpus. This thesis presents the following objectives.

2.1 Main Objectives

Our study presents the following general objectives:

- To evaluate the formal statistical quasi-distances for the comparison of two samples (corpuses) of the same historical text Nican Mopohua.
- To design a statistical model of quasi-distance that optimally detects historical linguistic variants and check its fit to statistical distributions by several GOF tests.

And also presents the following specific objectives:

- To test fit of empirical quasi-distances by GOF tests to the theoretically hypothetized statistical distributions, i.e. uniform and gamma.
- To detect the frequency of selected sets of words in the historical corpus.

To articulate the core of statistical reasoning Stigler (2016) explains the statistics under 7 principles or pillars that are: aggregation, information, likelihood, intercomparison, regression, design and residual. He tries to find approaches to analyze the data that produce

the information and therefore the knowledge. For this, we rely on the information theory of Shannon (1948) who published "A mathematical theory of communication" to improve the efficient transmission of information. Fisher (1925) also defined the information in statistics for the first time in his work on the Theory of Estimation. Lindley et al. (1956) considered that information is a statistical concept. Likewise, Kullback (1978) tells us that information theory is a branch of mathematical theory of probability and statistical mathematics.

This thesis studies objects of a linguistic nature which are analyzed by various aggregation operators defined on weighted linguistic information. Montero et al.(2010) tell us that the aggregation of information is fundamental for the decision, and this implies reduction of the original information while maintaining its format. Stehlík (2016) gives us an advantage of possibly a new discovery by considering proper underlying topology, indicating that aggregation is combination and fusion of several objects where the data topology allows the necessary variability for the information regularization.
The thesis solves problems of statistical inference, such as selecting an appropriate measure for multinomial or categorical data. But, we emphasize that the study is interesting in analyzing data that are presented through the quantifiable frequencies. The aim is to develop background for a social-linguistic study through the design of appropriate statistical method that optimally detects linguistic variants, by the frequency of letters, words and the presence of key words. We work with probabilities of co-occurrence of linguistic objects that is an important tool in statistics for studies of natural language (Lee, 2001).
Some measures of divergence between two probability distributions are developed in the thesis. Bigi (2003) recommends to use a measure of divergence called KL-divergence for the study of natural language. Kanamori and Sugiyama (2014b) warns us that there exist measures of discrepancy between two probability distributions where the difference between densities is based on the distances and/or the density ratio. Illert and Allison (2004) analyzes the aboriginal words, in this case the relative percentage of words. In addition, it considers that the change of the language is caused by the own evolution of the country . As a preprocessing step one can perform the stabilization of the variance since it is of vital importance to make comparisons. A type of transformation to stabilize the variance is proposed by Makri et al. (2017) working with binary sequence $(0-1)$ of Markov-dependent processes for the statistical characteristics of the historical corpus.
Bigi (2003) uses a statistical distribution for each corpus obtained from a training corpus. The comparison of the probability distribution of each letter or word of the corpus is made by the KL divergence, with a limited number of comparison terms for the corpus. The fact that the frequency of many terms in the corpus is zero is causing problems in the calculation of the KL

distance when the probabilities are estimated by frequencies of occurrences. KL-divergence is widely used in many language applications based on a statistical language model and information retrieval for identification topics.

Lee (2001) performed studies on the measures of distributional similarity by using the weighted average of the distance and found that the estimate based on similarity is more efficient (Dagan et al., 1999) and (Lee, 2000) . It indicates that the similarity between objects can be determined by the similarity of their corresponding vector characteristics, where these characteristics are numerical frequencies of co-occurrences. Also the estimators based on similarity is more robust than others, being therefore more efficient. Lee (2001) indicates that the skew divergence is the one that achieves the best performance and that is closest to a KL-divergence. Also, it analyzes many functions of similarity (or quasi-distance) such as:

KL divergence: $D(p \parallel q)$ is a measure of the distance between two probability functions. It runs into a problem of being undefined if there is a $x \in X$ so that $p(x) > 0$ but $q(x) = 0$. This makes the derived distributions inadequate for maximum likelihood estimators, which would assign probabilities of zero to the co-occurrences that do not appear in the data of one of the given two corpuses (Essen and Steinbiss, 1992). In addition, it tells us that one option is to use smoothed estimators so $q(x)$ is not zero for all x. Another option is to use approximations of the KL divergence that does not require p to be absolutely continuous with respect to q. Let us consider Jensen-Shannon divergence (Lin, 1991) (Grosse et al., 2002) and the Skew divergence (Lee, 2001). Jensen-Shannon is symmetric and considers the KL divergence between p and q and the average of p and q. On the other hand, the asymmetric Skew divergence simply smoothes one of the distributions by mixing it to a degree determined by the parameter α with the other distribution.

Other measures of distributional similarity are: Euclidean distance, cosine and distance L_1 (or Manhattan). It also includes "probability confusion", which estimates the substitutability of two given words, based on conditional and marginal probabilities.

Table 1.1 displays some similarity functions for probability distributions used in this thesis.

Table 2.1 Similarity functions for probability distributions

KL DIVERGENCE	$D(p \parallel q)$	$= \sum_x p(x)(logp(x) - logq(x))$
JENSEN-SHANNON	$JS(p,q)$	$= \frac{1}{2}[D(p \parallel avg(p,q)) + D(q \parallel avg(p,q))]$
SKEW DIVERGENCE	$S_\alpha(p,q)$	$= D(q \parallel \alpha p + (1-\alpha)q)$
EUCLIDEAN	$euc(p,q)$	$= (\sum_x (p(x) - q(x))^2)^{\frac{1}{2}}$
COSINE	$cos(p,q)$	$= \sum_x p(x)q(x) / \sqrt{\sum_x p(x)^2 \sum_x q(x)^2}$
L_1	$L_1(p,q)$	$= \sum_x \lvert p(x) - q(x) \rvert$
CONFUSION	$conf(p,q,P(x')) =$	$= P(y') \sum_x p(x)q(x)/P(x)$

Kanamori and Sugiyama (2014a) tells us about the existence of two measures of discrepancy, that is, distances and divergences, here the intersection of them is the L_1-distance. L_1 distance can be estimated by the measurement of distances based on the difference of densities and the measure of the divergence can be estimated based on the density ratio. In addition, they find that estimation of the distance L_1 is more robust than estimation of the density ratio.
Also the difference in densities $p - q$ is used to calculate the distance L_s between two probability densities:

$$d_s(p,q) = \left(\int |p(x) - q(x)|^s dx \right)^{\frac{1}{s}}$$

where $s \geq 1$.

Given that our information is linguistic and assuming that the larger amount of information can reduce the uncertainty, all the data of the written compilation or corpus was analyzed. Our corpus is a historical jewel called Nican Mopohua (Lasso,1649) based on the written native language of Nahuatl. The study focuses on identifying and locating the relative position of each letter or word. It also focuses on the semantics of the ancestral or native language (Nahuatl) and finally Stehlík (2016)'s Mereological and topological notions of connection, part, interior, and complement wich are central to spatial reasoning and to the semantics of natural language expressions. In this chapter we will review some preliminaries for better orientation in the thesis.

2.2 Aggregation in topological spaces: topological closeness of linguistic objects

Morphemes, roots and affixes are very important neighbouring structures from topological point of view of any natural language. According to Stehlík (2016) an aggregation function is a function that is applied to a set of values and returns a single, aggregated value. In what follows we provide definition of aggregation function obtained by direct topologization of the metric aggregation function introduced by (Grabisch et al., 2009)

Definition (see Stehlík, 2016) $(X,T,\leq)$ is a topological space with partial ordering $\leq$, and T is the family of the open subsets of X. An aggregation function in X^2 is a function $A : X^2 \to X$ such that

(i) it is non-decreasing in each variable, and

(ii) for each $u \in X$ there is $(x_1,x_2),(y_1,y_2) \in X^2$ such that $A(x_1,x_2) \leq u \leq A(y_1,y_2)$

Here a relation $\leq \subset X \times X$ is called a partial ordering if it is reflexive, antisymmetric and transitive.

Remark 1 (see Stehlík, 2016) Notice that we do not need X to be the topological space to define an aggregation function in the definition. If X is a bounded chain, with top element 1 and bottom element 0, then (ii) is equivalent to the idempotency of 0 and 1. So, the definition 1 can easily be generalized to define an $n-ary$ aggregation function.

Remark 2 (see Stehlík, 2016) Notice that topological aggregation is done in the context of general mappings $A : X^2 \to X$, where X is a topological space, sometimes with additional properties.

Now presenting the formal definition,we will recall some topological concepts. If p is an open cover of a topological space (X,T), and $x \in X$, the star of the point x is defined as

$$st_p(x) = \bigcup_{V \in p, x \in V} V$$

By induction, having $st_p^1(x)$ we can define $n \geq 2$ the $n-star$ of x by

$$st_p^n(x) = \bigcup_{V \in p, V \cap st_p^{n-1}(x) \neq \phi} V$$

Where the star of the point is a convenient relaxation of the metric concept.

A system of open sets $A \subset T$ is called open cover of space X iff $\bigcup_{M \in A} M \supset X$

2.2.1 Examples on topological proximity of Náhuatl words

The Náhuatl as the agglutinating language presents words consisting of lexemes (root) and affixes (prefixes, suffixes and infixes) as can be seen in the following words:

a) **Quimo**lhuili (told him), **quimo**nanquili (he answered), **quimo**nonochilia (comunicates), **quimo**nahuatili (orders him), **quimo**tlatlauhtili (begs),**quimo**lhuilia (he tells), **quimo**ttili (he saw it) and **quimo**nanquilili (he answered).

b) **iyo**llo (your heart), i**iyo**tzin (his venerable breath), ni**iyo** (my essence), mi**iyo**tzin (your venerable breath), totecu**iyo** (our lord) and notecu**iyo**é (oh my lord).

c) **Teo**tl (God), **teo**yotl (divinity), **teo**pixcatlatoani (bishop), to**teo**pixcahuan (our priest), no**teo**cal (he answered) and i**teo**caltzin (her venerable temple).

d) **Oqui**cac (he hear)and **oqui**ttac (he saw).

e) **Cihuapill**i (maid) and **cihuapill**é (princess).

f) **Tlacatl** (man) and **tlacatl**é (mistress of humanity).

g) **Xochitl** (flower) and Tlazo**xochitl** (beautiful flowers).

Next, we present the figures of the topological neighbors that illustrate the lexemes and affixes of the 7 words:

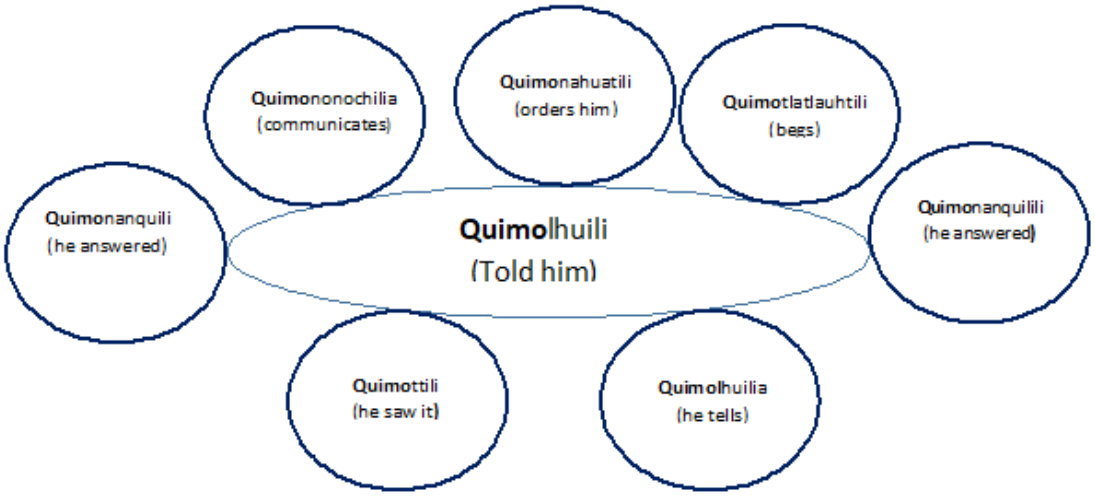


Fig. 2.1 Nearest neighbor of the word: **Quimo**lhuili (told him)

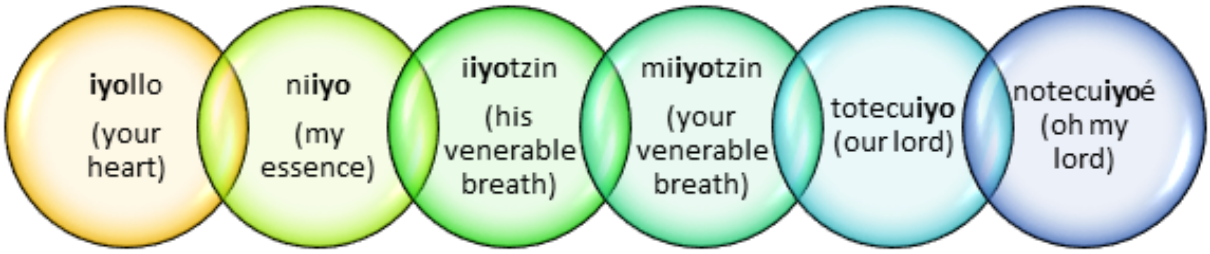


Fig. 2.2 Nearest neighbor of the word: **Iyo**llo (your heart)

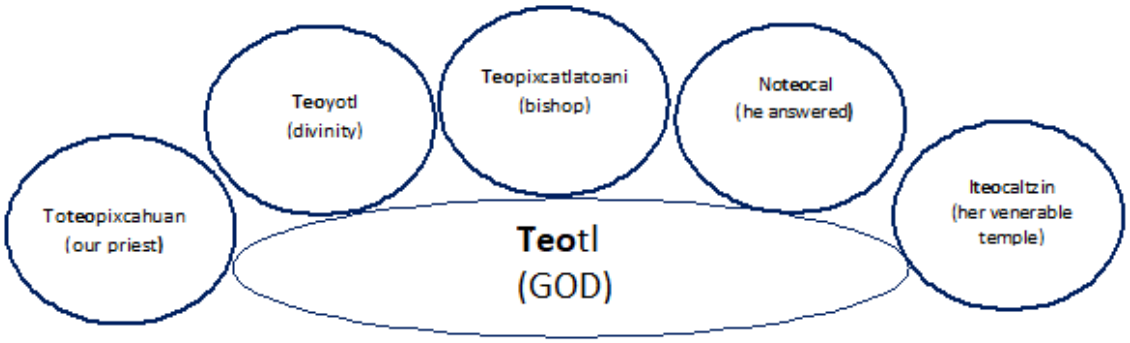


Fig. 2.3 Nearest neighbor of the word: **Teo**tl (God)

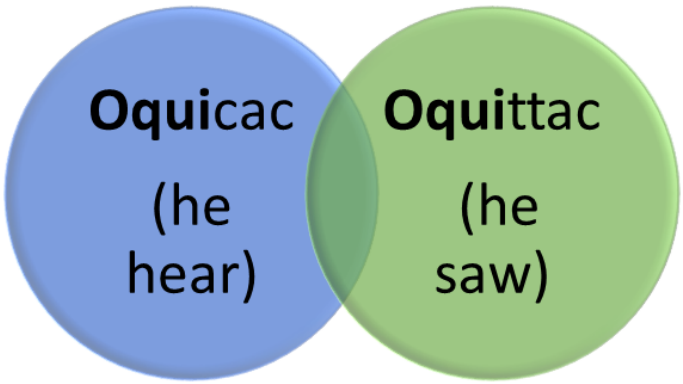


Fig. 2.4 Nearest neighbor of the word: **Oqui**cac (he hear)

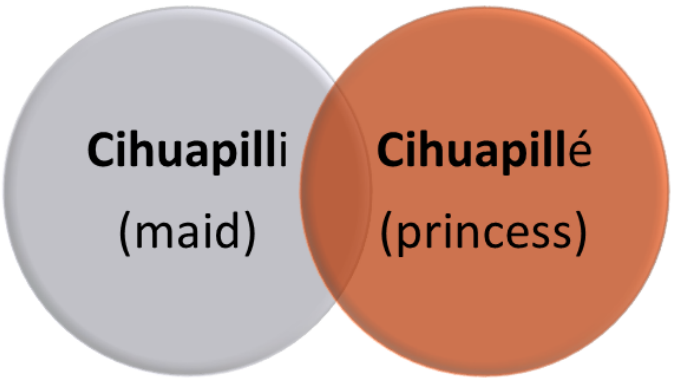


Fig. 2.5 Nearest neighbor of the word: **Cihuapill**i (maid)

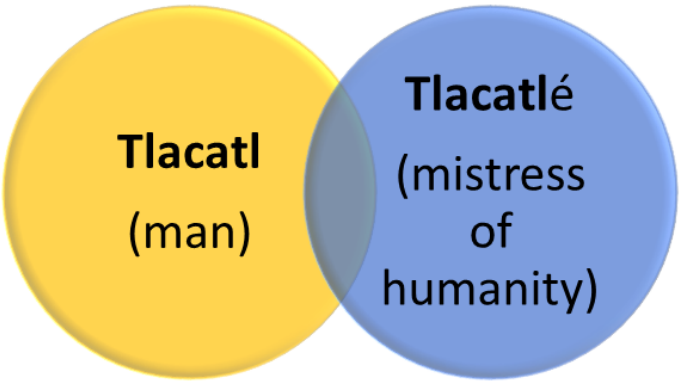


Fig. 2.6 Nearest neighbor of the word: **Tlacatl** (man)

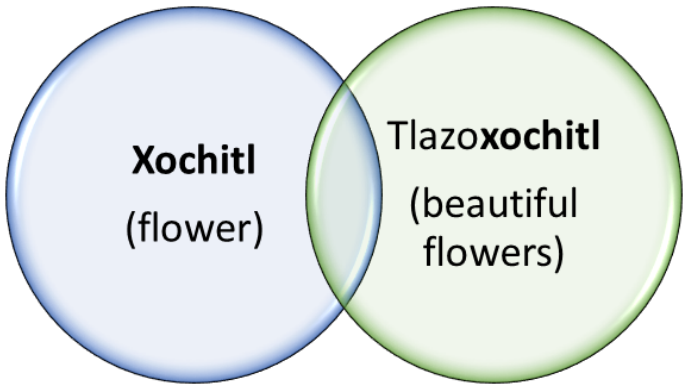


Fig. 2.7 Nearest neighbor of the word: **Xochitl** (flower)

2.3 Nican Mopohua

The Nican Mopohua, historical Nahuatl text, meaning "Here is narrated" or "Here is related", is considered as the collection of written expression or corpus that allowed us to study and analyze the variations of the Nahuatl language. The original of the Nican Mopohua was written on paper made with maguey pulp, like the ancient Aztec codices. The writer used the Latin characters recognized as those learned by the natives in the first stage of their conversion to Christianity and consequent incorporation into European culture (Rojas, 1978). This document has figurative words which explains the profound meaning of the life in which the Aztec culture was founded. The Nahuatl used in this document is so elegant, subtle and refined. A characteristic of language is that it is of a blended nature that imposes an instantaneous mental inspection on the meaning of each particle and on the contextual sense of its total construction. That makes it rich in meanings and connotations.
In one of the most important editions of the study, the historic document is contained in a larger book entitled Huei tlamahuizoltica omonexiti in Ilhuicac tlatohcacihuapilli Santa Maria Totlazonantzin Guadalupe in nican Huei altepenáhuac Mexico itocayocan Tepeyácac (1649) which means "a great miracle appeared the Heavenly Queen, our precious Mother Mary of Guadalupe, near the great Altepetl of Mexico, where they call Tepeyacac ", consists of 36 pages where, apart from the Nican Mopohua, another text called Nican Motecpana appears. This was published by the bachelor Lasso De La Vega (1649), vicar of the Tepeyac chapel. This document of identity and Mexican national integrity allowed us to analyze the morphology of the Nahuatl language that was written by Antonio Valeriano in 1556 and where the appearance of the Virgin of Guadalupe to Juan Diego is narrated on the hill of Tepeyac in December 1531. Our research analyzes the writing of the language. In the next we digress to the elementary notions of Nahuatl language.

Ortiz (1990) indicates that the human being is essentially sociable therefore communication is essential to express ideas. It also informs us that there were several indigenous languages of the XVI century that were discovered by the conquerors and that these languages had a high degree of development in human groups of great culture such as Nahuatl. Nahuatl language is beautiful, very extensive, flexible, rhythmic, poetic and highly expressive. In addition, in their wisdom they were retailers even to put the names and that they did it seeing the characteristics of the places and the one of the inhabitants with the most human sense and with a presence of a poetic way of speaking.
The native speakers did not use letters, the writing was the hieroglyphics under a need to perpetuate their deeds to future generations and they used only the educated classes as: priests,

military and administrative chiefs, chief merchants, fortune tellers and scribes. These hieroglyphics were found by the conquerors in the early sixteenth century, in a mixed evolutionary period, in which without leaving the figurative (rough representation of the object), nor the symbolic (representation of the object and stylization), were used ideological (representation of ideas abstract or concrete by conventional signs) and phonetics were reached (representing a syllable or a letter that serves to express ideas independent of the drawing). Thus, the missionaries then introduced the use of the letters of Castilian for the writing of Nahuatl, using the alphabetic signs that best represented the sounds they heard. But in that century the Castilian language lacked orthographic rules since only in 1713 the first Spanish dictionary was published, and all educated people used Latin. The orthographic anarchy is reflected in the writing of the Nahuatl with Castilian letters, for that reason the Latin was resorted to be used, but this also caused confusion in the spelling. Other difficulties are that in Spanish most syllables are direct, in Nahuatl they are inverse since the vowels naturally tend to join in the following consonant and to separate a vowel from the consonant that follows it. This is done in the pronunciation a little aspiration, that the grammarians of the XVI century call saltillo and in the current writing it has been coming to use an "h". For example, "tlacohtli" (server).

When fixing in letters of the Spanish alphabet in the sounds of the Nahuatl language there were sounds that did not have a sign of writing in Spanish. It is the case of the sound "sh" represented with the letter "x". But, at the end of the 16th century, until the seventeenth century when the "x" took the sound "j". The exchange of "s" and "x" was also frequent. But, in 1815 the language authorities determined that it should only have the "cs" sound.
Ortiz (1990) defines the Nahuatl language as a flexible, poetic and highly expressive language that has the following characteristics:

a) Polysyllabic, because most words are multi-syllabic.

b) Binders, where the constituent elements of words lose something of themselves when grouped.

c) Polysynthetic, because for word translation into Spanish it uses sometimes a whole phrase.

Where, also the compound words originate from a nucleus, called semantema and that can consist of one or two elements and one or more particles called morphemes, which complete or modify the meaning. If these morphemes are located at the beginning they are called

pre-semantics or prefixes and if they go to the end postsemantic or endings.
Something that distinguishes Nahuatl from other indigenous languages is that the words of more than one syllable carry the prosodic accent in the penultimate syllable, grave or flat, and if syllables are added or removed at the end, the accent changes to remain flat. It is also necessary to emphasize that in the writing the accent is not used. Another important observation in the language is when the word that wants to express esteem and respect is added to the reverential "tzin". For example: "itahtzin" (the dear father).
When the conquest arrived, the use of letters prevailed and the missionaries fixed by ear, with letters, the sounds they thought they perceived. Also, Castilian and Nahuatl of the sixteenth century has a different script to the current one. It is therefore, that there are no firm bases in the language. But Ortiz (1990) recognizes 24 letters so for our research analysis we will use the classic Nahuatl, recognizing 21 letters of the Nahuatl alphabet.

The intermediate sounds e-i, o-u, have given rise to many confusions by the imposition of Castilian phonetics to the native language. The line that is presented on some vowels is to differentiate the long sounds from the short ones, since this also changes the meaning of the word. Therefore, it is a language of subflexion, because it does not have all the inflections of other languages. Some peculiarities of the Nahuatl language.

Nouns:

The noun presents four grammatical numbers being: Singular (for names of inanimate things, unless they are personified), dual, plural and collective. For a reverential, estimative, diminutive mode of affection, "tzintli" is used at the end. For example "icnohuahcanantzintli" (revered mother).

Verb:

It is the verb itself, it is the third singular person of the indicative present, which is translated into the infinitive. Pronominal indicators are required for first and second persons, except for the complement in reflexive verbs. Also, that transitive verbs require complement. And for the past tense, the verbs are prefixed with the words "o". Example "otlahtoh" (he / she speaks).
The verb is classified by its meaning, function, structure and complexity.

Participles:

Like the Spanish, the participles derive from verbs and also present the two classes: active (execute the action) and passive (receive the action); the first ones are: temporary (when you execute the action occasionally) and constants (when you execute the action by habit or trade) and the seconds, are formed with the particle of indefinite thing complement.

Adjectives:

There are the primitive adjectives and derivatives. As a rule, adjectives and adjectival particles go before the noun or substantival part and are invariable before gender and in terms of number.

On the other hand, in the field of numbers, it presents a vigesimal base numbering.

In addition, recognizing that the writing is subsequent to the word and originates when the said thing is fixed. In the Nahuatl language, using the alphabet, different sounds are given to the same letter. In the sixteenth century ç (cedilla) is used, the "u" and the "v" are interchangeable, as well as the "v" and the "b", and the meaning of some words is different from now. Also, without using special signs for sounds that Spanish does not have, writing is complex, there are cases where a word is given two different meanings, for not differentiating the roots. In addition, all words have a grave or flat accent, syllables are often inverse, and when two elements are joined, a letter is sometimes elided. It is also important to indicate when a vowel is long or short.

The presence of religious inclinations, the supreme deity, the dual divinity and its manifestation in natural phenomena makes the indigenous people sensitive and inspiring. Likewise, they consider that the basic cell of society is marriage, contributing in this to the development of poetry in the Nahuatl language.

Poetry:

It was a way to transmit to the new generations their culture under a literary expression that is given in prose or poetry. Nahuatl-speaking people were also recognized as the "Nahuatlacah", because they are of nature poets, since the language presents clear and harmonious sound, softness, sweetness, rhythmic accentuation, rhythmicity, and the possibility of forming words with defined rules. Poetry was transmitted by constant oral repetitions of memory. In addition, that the song was conditioned in the genre of poetry by dance: music, song and dance were normally united.

Another characteristic of the Nahuatl poetry is that it is collective, not individual and is impregnated with a religious sense that celebrated the divinity, with a brief tendency in the poems but of great transcendence. Being a closed culture where the paradigm of beauty expresses it: the flowers, the birds of beautiful plumage and the precious stones. Determining its authenticity. For all this, Nahuatl poetry has the right to survive and to be appreciated by humanity. The Náhuas poets took several subjects such as: the religious, the warrior or heroic, the philosophical and the personal. The poetic technique is reduced to the short poem and two times: the first enunciates the thought and the second is completed and both are

linked by a ritornelo or refrain.
There is a proof of the existence of groups or schools of poets, these were called "cohuáyötl, icniúhyötl" which were societies or fraternities formed to cultivate song and collective joy. The Nahuatl literary production said his way of understanding the world and life, where the softness of the language used to express thoughts and feelings stands out.

Our research conducted from design and realization of two experiments with the information base of the Nican Mopohua (historical corpus) which is made up of 218 paragraphs. In addition, a comparative analysis of letter and key word frequencies was carried out in the historical corpus of Lasso De La Vega (1649) and Rojas (1978). It is also necessary to indicate that the paragraphs of 88-93 are not present in the corpus of Luis Lasso de la Vega. This omitted the 6 paragraphs referring to "the third appearance of the virgin to Juan Diego, telling him to return tomorrow to take the message to the bishop and where he communicates again that his effort will be rewarded." But, Rojas (1978) reports that they had already been written by other researchers such as Fr. Miguel Sánchez, and later by Becerra Tanco himself who used the original source. It is then where Rojas uses the Castilian text of that passage and translates it into the language of Valeriano (see Rojas (1978)). This event can adulterate the veracity of the historical document with the 218 paragraphs.
Therefore, only the remaining 212 and common paragraphs for both corpus were analyzed. It is also necessary to emphasize that our study is based on the analysis of classical Nahuatl, since the historical document (Nican Mopohua) is written in the language spoken by the Mexicas (Aztecs) of Mexico-Tenochtitlan (the center of what is now Mexico City). With the writing established by the Spanish friars who elaborated an alphabet based on Latin-Spanish letters (before the arrival of Europeans the language had a writing with a partially ideographic system)

In the first experiment **A**, the frequency of letters and key words in 9 paragraphs of 118-126 were analyzed. These paragraphs were selected because they counted the largest number of key words and represented the paragraphs of greatest message of the deepest feeling of love, peace and consolation of the Virgin. Also, the Nahuatl alphabet was set with 20 letters that were the following: a, b, c, d, e, g, h, i, j, l, m, n, o, p, q, t, u , x, y, z. Also, this was obtained by deducting and spelling the historical corpus decomposing and arriving at the frequency of each of the letters for the nine paragraphs mentioned above. The translations of the key words were carried out with the GDN digital dictionary (Gran Nahuatl dictionary) with the URL http://www.gdn.unam.mx/termino/search. Also, the Kullback-Leibler divergence measure was calculated to demonstrate the asymmetry between the two probability

functions of selected keywords (see also Stehlík and Pari (2018) Discussion on Pigoli et al. (2018) "The statistical analysis of acoustic phonetic data: exploring differences between spoken Romance languages", *J. R. Statist. Soc. C* Applied Stat. 67, Part 4.)

The following Table 2.2 displays analysis of experiment (**A**). First we computed relative frequency of each letter, with the absolute frequency divided by the total number of letters (1027 letters in the 9 paragraphs of Lasso corpus and 1046 letters in the Rojas corpus). For example the letter "a" appears 99 times in 9 paragraphs (118-126), this divided by the total of letters 1027 result 0.0964 and that multiplying by 100 we will have the percentage frequency. In the same way we proceed to find the relative frequencies for Rojas corpus. Next, the KL-divergence was calculated with the two relative frequencies of the corpus. Where the probability of$P$ is in the corpus I of Lasso and the probability Q is in the corpus II of Rojas. Then the quasi - distance value is calculated (see table 2.1 KL divergence) for each letter of the alphabet and finally the sum of the 20 quantities of the letters of the alphabet gives us as a result the value of the quasi-distance $D(P||Q)$. Similarly we obtain $D(Q||P)$ to check its asymmetric property. The different values 0.0005202 of $D(P||Q)$ and 0.0005891 of $D(Q||P)$ indicate that the two probability density functions are different. Nevertheless the numerical difference looks to be small, we shall point out that more careful analysis confirms asymmetry, underlined by conjecture that since one corpus evolved from another one, the underlying random walk model is not revertible. Therefore, the values found in the first letter frequency counting experiment show some similarity between the pdfs of intuition and deduction of the Nahuatl alphabet. From this point of view we can consider occurrences of some words to be affected by sort of omissing mechanism of linguistic ambiguity.

Table 2.2 **A**: Corpus I (Lasso-1649) Vs Corpus II (Rojas-1978)

Letters	Frequency		$p_{letter}.\log\left(\frac{p_{letter}}{q_{letter}}\right)$	$q_{letter}.\log\left(\frac{q_{letter}}{p_{letter}}\right)$
	Corpus I(%)	Corpus II(%)		
a	9.64	10.99	-0.00550442	0.006277879
b	0.10	0.10	7.75194E-06	-7.61113E-06
c	10.13	9.85	0.001231125	-0.001197139
d	0.10	0.10	7.75194E-06	-7.61113E-06
e	3.99	4.21	-0.000906534	0.000955195
g	0.10	0.10	7.75194E-06	-7.61113E-06
h	4.87	4.59	0.001250731	-0.001178891
i	16.75	16.63	0.000492459	-0.000489136
j	0.10	0.10	-1.12663E-05	1.15704E-05
l	6.43	6.21	0.000937741	-0.000906757
m	5.06	4.97	0.000403101	-0.000395779
n	9.93	9.75	0.000790698	-0.000776335
o	8.37	8.41	-0.000169402	0.000170193
p	2.04	2.01	0.000162791	-0.000159834
q	2.04	1.91	0.000596068	-0.000557372
t	7.40	7.36	0.000169028	-0.000168142
u	6.52	6.21	0.001378014	-0.001312595
x	1.95	1.91	0.000155039	-0.000152223
y	1.85	1.91	-0.000264837	0.000273712
z	2.53	2.58	-0.000213397	0.000217579
TOTAL			$D(P\|\|Q)$= 0.0005202	$D(Q\|\|P)$= 0.0005891

The second experiment (**B**) was carried out with the collaboration of Temachtiani (native language teacher) Mary, who serves a professor at the House of Culture of Azcapotzalco of the D.F. of Mexico and the bilingual native teachers Victoriano de la Cruz Cruz and Mireille Nallely Sánchez Olivares. The first serves as professor of the Department of Studies in Indigenous Languages of the CUCSH (University of Guadalajará, Mexico) while the second is professor of the National Autonomous University of Mexico (UNAM). The support received from these great native professionals and Mexican educational scholars was the correction of the Nahuatl alphabet and the translation of the 111 key words. For the analysis of letter and word frequencies, we worked with all the paragraphs (218) of the historical corpus. Based on Garibay (1940) the classic Nahuatl alphabet consists of 20 letters, some

of them digraphs, as follows: a, c, ch, e, h, i, l, ll, m, n, o, p, q, t, tl, tz, u, x, y, z. In the experiment we evaluated empirical version of the Kullback-Leibler divergence measure, in particular also to demonstrate the asymmetry between the two empirical probability functions. We obtained the following values edited in the following Table 2.3.

Table 2.3 **B**: Corpus I (Lasso-1649) Vs Corpus II (Rojas-1978)

Letters	Frequency		$p_{letter}.\log\left(\frac{p_{letter}}{q_{letter}}\right)$	$q_{letter}.\log\left(\frac{q_{letter}}{p_{letter}}\right)$
	Corpus I	Corpus II		
a	0.1083	0.1067	0.0007267	-0.0007156
c	0.0630	0.0685	-0.0022672	0.0024631
ch	0.0127	0.0126	0.0000406	-0.0000403
e	0.0505	0.0500	0.0002209	-0.0002187
h	0.0538	0.0530	0.0003481	-0.0003429
i	0.1741	0.1715	0.0011481	-0.0011308
l	0.0600	0.0592	0.0003242	-0.0003202
ll	0.0049	0.0049	0.0000015	-0.0000015
m	0.0358	0.0356	0.0001131	-0.0001123
n	0.1025	0.1014	0.0005040	-0.0004983
o	0.0765	0.0762	0.0001440	-0.0001434
p	0.0200	0.0203	-0.0001318	0.0001339
q	0.0291	0.0279	0.0005317	-0.0005098
t	0.0417	0.0415	0.0001072	-0.0001066
tl	0.0248	0.0248	-0.0000094	0.0000094
tz	0.0132	0.0130	0.0000887	-0.0000874
u	0.0746	0.0746	0.0000163	-0.0000163
x	0.0096	0.0096	-0.0000208	0.0000209
y	0.0173	0.0169	0.0002067	-0.0002011
z	0.0275	0.0320	-0.0018118	0.0021087
TOTAL			$D(P\|\|Q)$= 0.0002809	$D(Q\|\|P)$= 0.0002908

The Table 2.3 shows in the second and third column the relative frequencies of each of the 20 letters of the alphabet. That is, the absolute frequency of each letter between the number of letters in the entire corpus (total 218 paragraphs) identifying that the Lasso corpus has 20,408 letters and the Rojas corpus comprises 21,035 letters. And in the fourth and fifth column are the values obtained from the calculation of the KL divergence of each of the

letters. Finally, in the last row of the table this the sum of all the values of quasi-distance. The values 0.0002809 of the $D_{KL}(P||Q)$ and the 0.0002908 of the $D_{KL}(Q||P)$ indicates that the two probability density functions are different. where there is a proximity in the empirical frequencies of the keywords and letters of the corpus I (1649) and corpus II (1978) articulates similarity of a classic Nahuatl alphabet.

2.4 Results of 111 keywords and gamma GOF tests

The data is composed of 111 key words of all Nican mopohua corpus. These are expressed as relative frequency values obtained from the comparison of two corpuses in relation to their frequency of occurrence. Then the similarity measures were applied: KL divergence, Jensen- Shannon, Skew divergence, Cosine, L_1 and Confusion. The results indicate that the KL-divergence yields better results although its values are close to zero, that is, the KL-divergence between the words is minimal, even though the corpuses are distant by time of 329 years, indicating that the keywords have not much changed in spelling and semantics. Natural question arrives, whether the generalized gamma distribution (ggd) limit of KL divergence quasi-distance is valid here, in particular, whether gamma submodel of ggd can be sufficient. To clarify this question, we applied selected gamma GOF-tests, including a new GOF-test for gamma distribution developed by Villaseñor and González-Estrada (2015) who proposes the estimator V_n* (see equation 1.6, page 7) that is based on the ratio of two variance estimators with the asymptotic null distribution . The calculation is conducted by the *gamma_test* function of the package textit fitdistrplus (see appendix B). We also applied battery of other goodness-of-fit tests, like Log likelihood (logL), Akaike information criterion (AIC), Bayesian information criterion (BIC), chi square test (χ^2), Anderson Darling test (AD) and the Kolmogorov Smirnov test (KS). These GOF-tests were calculated with the *fit.cont* function of the *rriskDistributions* package, developed by Belgorodski et al. (2017) (see appendix B). We compare the values of each of the GOF-tests and found some discrepancy with the estimator (V_n*) of Villaseñor and González-Estrada (2015) in some quasi-distance functions. These discrepancies can be related to slow convergence rate of empirical statistics to theoretical asymptotical limiting distribution. The majority of the GOF-tests concluded that the data cannot be rejected to follow a gamma distribution with the strict exception of the Cosine and Confusion functions. The latter ones follow Beta and Cauchy distributions respectively.

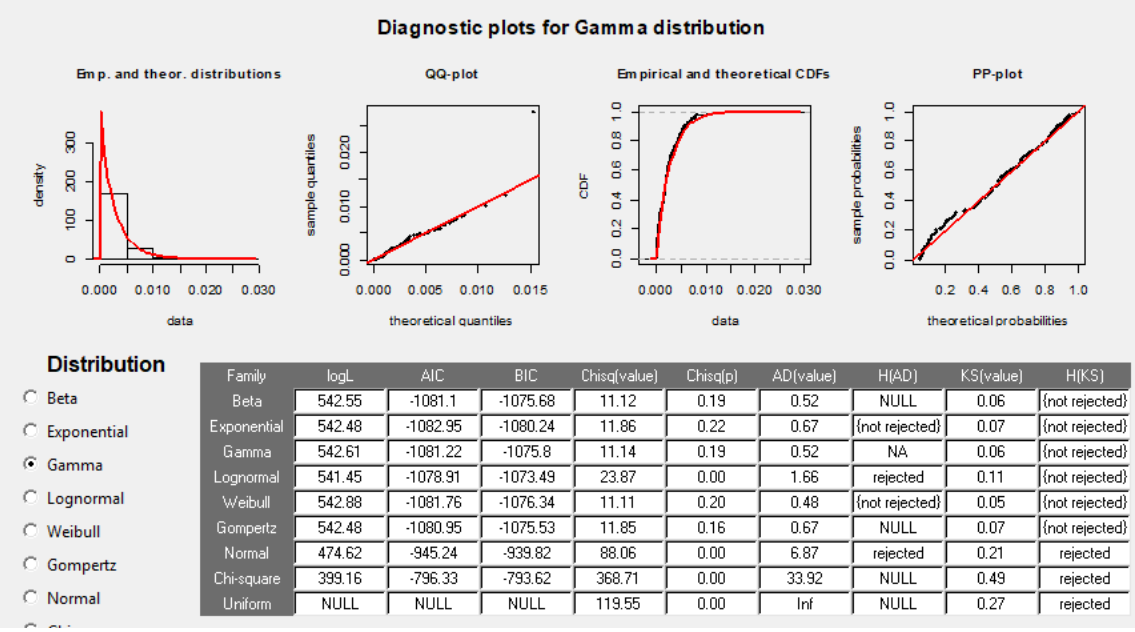
The figures 2.8 to 2.17 display the results of the first experiment with 111 key words that were selected for the greatest number of frequencies presented in the historical corpus. It should be noted that Náhuatl is an agglutinative language (words are formed by joining independent monemes) which is characteristics common to many native languages.

a) KL Divergence D(p||q):
D($p||q$): GOF-test for the Gamma distribution

$$V_n^* = 3.1271, \quad V_n = 1.327099, \; p - value = 0.02702$$

The goodness of fit test for H_0 based on the ratio of two estimators of the population variance (V_n) indicating that the null hypothesis is not rejected with a level of significance of 0.01 .
In addition, we can say that the gamma distribution has the following parameters obtained from R-package *rriskDistributions*:

$$Shape = 0.9419465 \qquad Rate = 339.4442287$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	542.55	-1081.1	-1075.68	11.12	0.19	0.52	NULL	0.06	{not rejected}
Exponential	542.48	-1082.95	-1080.24	11.86	0.22	0.67	{not rejected}	0.07	{not rejected}
Gamma	542.61	-1081.22	-1075.8	11.14	0.19	0.52	NA	0.06	{not rejected}
Lognormal	541.45	-1078.91	-1073.49	23.87	0.00	1.66	rejected	0.11	{not rejected}
Weibull	542.88	-1081.76	-1076.34	11.11	0.20	0.48	{not rejected}	0.05	{not rejected}
Gompertz	542.48	-1080.95	-1075.53	11.85	0.16	0.67	NULL	0.07	{not rejected}
Normal	474.62	-945.24	-939.82	88.06	0.00	6.87	rejected	0.21	rejected
Chi-square	399.16	-796.33	-793.62	368.71	0.00	33.92	NULL	0.49	rejected
Uniform	NULL	NULL	NULL	119.55	0.00	Inf	NULL	0.27	rejected

Fig. 2.8 gamma distribution: $D(p||q)$

Figure 2.8 contains the analysis of the similarity function of the Kullback-Leibler divergence $D(p||q)$, where the first graph presents the density of the gamma distribution.

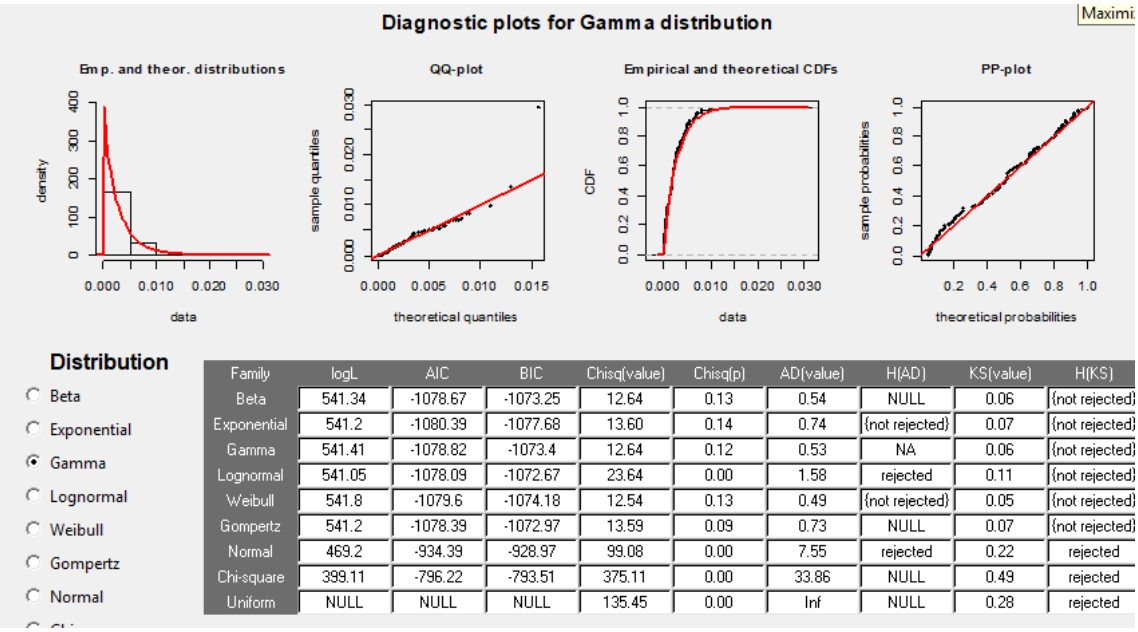
On the second graph Q-Q plot visually determines that most data approaches a straight line although it presents outliers. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data fits appropriately to the distribution and the last graph determines that the data are distributed or exhibit a gamma behavior although some discrepancy between the red line with the foreground in 0.3 approximately. Also below it presents a table that reflects the numerical values of the different goodness of fit tests: logL that presents a high value of 542.61. The AIC has a minimum value of -1081.22. BIC also has a low value of -1075.8. The smaller chi square value of 11.14 tells us that it fits. Also the lower AD value of 0.52 and the KS value of 0.06. Consequently, all the values of the different GOF-tests assert that the distribution of the data fits the probabilistic gamma model.

D($q||p$): GOF-test for the gamma distribution

$$V_n* = 3.7413, \quad Vn = 1.394387, \; p - value = 0.008157$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (V_n) indicating that the null hypothesis is rejected with a level of significance of 0.01.
Under assumption that the distribution of D($q||p$) is gamma, we have the following estimators:

$$Shape = 0.9269453 \qquad Rate = 330.1916554$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	541.34	-1078.67	-1073.25	12.64	0.13	0.54	NULL	0.06	{not rejected}
Exponential	541.2	-1080.39	-1077.68	13.60	0.14	0.74	{not rejected}	0.07	{not rejected}
Gamma	541.41	-1078.82	-1073.4	12.64	0.12	0.53	NA	0.06	{not rejected}
Lognormal	541.05	-1078.09	-1072.67	23.64	0.00	1.58	rejected	0.11	{not rejected}
Weibull	541.8	-1079.6	-1074.18	12.54	0.13	0.49	{not rejected}	0.05	{not rejected}
Gompertz	541.2	-1078.39	-1072.97	13.59	0.09	0.73	NULL	0.07	{not rejected}
Normal	469.2	-934.39	-928.97	99.08	0.00	7.55	rejected	0.22	rejected
Chi-square	399.11	-796.22	-793.51	375.11	0.00	33.86	NULL	0.49	rejected
Uniform	NULL	NULL	NULL	135.45	0.00	Inf	NULL	0.28	rejected

Fig. 2.9 gamma distribution: $D(q||p)$

Figure 2.9 shows the exploratory results for Kullback-Leibler divergence D($q||p$), the first graph represents the density of the gamma distribution. On the second graph Q-Q plot reveals that most of the data approaches a straight line even though it presents data outliers. On the third graph it is observed that the stepped line closely follows the adjusted distribution line presenting an adequate adjustment and the last graph determines that the data sets are distributed or exhibit a gamma behavior, although some discrepancy is observed between the red line with the first plane in 0.3 approximately. A table reflects the numerical values of the different goodness of fit tests: log likelihood (logL) that presents a high value of 541.41. The AIC has a minimum value of -1078.82. BIC also has a low value of -1073.4. The smaller chi square value of 12.64 tells us that it fits. Also the lower AD value of 0.53 and the KS value of 0.06. All the values of the different GOF-tests assert that the distribution of the D($q||p$) fits the probabilistic gamma model.

Here V_n^* stands for the statistical GOF-test critical values maximizing the asymptotic type I error probability on the space of the shape parameter and V_n is a test statistic based on the ratio of two estimators the unbiased sample variance (S_n^2) and the variance of gamma distribution ($\breve{\sigma}_n^2$). The p-value shows us that the D($p||q$) follow approximately gamma so that the null hypothesis is not rejected (under a level of significance of 0.01). On the other hand, the p-value for $D(q||p)$ shows us that the $D(q||p)$ does not follow a gamma and the

null hypothesis is rejected.

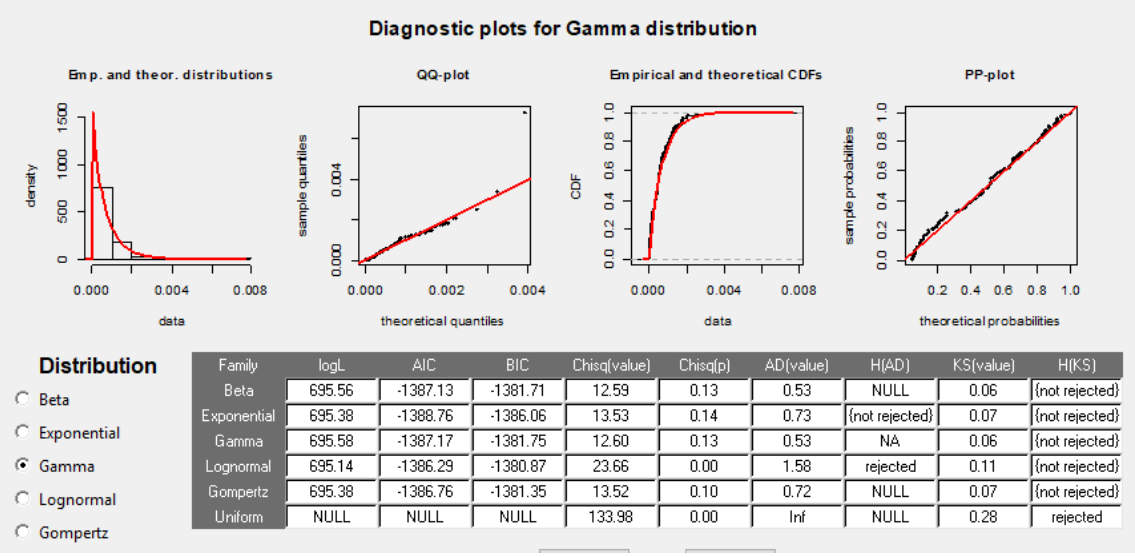
We also analyze the distribution of quasi-distances by the battery of GOF-tests (see Appendix B for R package *rriskDistributions*). We obtained the results shown in figures 2.8 and 2.17 where the collection of GOF-tests contains Log likelihood (logL), Akaike information criterion (AIC), Bayesian information criterion (BIC), chi square test (χ^2), Anderson Darling test (AD) and the Kolmogorov Smirnov test (KS) with the different distribution families indicate in a hierarchical way the best type of distribution that comes from the data. Seeing the values of the different GOF-tests we can affirm the intuition indicating that $D(p||q)$ and $D(q||p)$ both present a behavior approximately Gamma. The value of V_n* and its respective p-value in the measure $D(q||p)$ would be without effect since the rest of the goodness-of-fit tests determine it and also due to slow convergence of the empirical test statistics to limiting distribution.

b) Jensen-Shannon (JS):
JS(p,q)**:** GOF-test for the Gamma distribution

$$V_n^* = 3.683, Vn = 1.387774,\ p-value = 0.009206$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is rejected with a level of significance of 0.01.
The estimated parameters for gamma distribution are:

$$shape = 0.9290395 \qquad rate = 1327.2426465$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	695.56	-1387.13	-1381.71	12.59	0.13	0.53	NULL	0.06	{not rejected}
Exponential	695.38	-1388.76	-1386.06	13.53	0.14	0.73	{not rejected}	0.07	{not rejected}
Gamma	695.58	-1387.17	-1381.75	12.60	0.13	0.53	NA	0.06	{not rejected}
Lognormal	695.14	-1386.29	-1380.87	23.66	0.00	1.58	rejected	0.11	{not rejected}
Gompertz	695.38	-1386.76	-1381.35	13.52	0.10	0.72	NULL	0.07	{not rejected}
Uniform	NULL	NULL	NULL	133.98	0.00	Inf	NULL	0.28	rejected

Fig. 2.10 gamma distribution: $JS(p,q)$

Figure 2.10 shows the analysis of the Jensen-Shannon similarity function $JS(p,q)$. The first graph that shows the density of the gamma distribution. On the second graph, the Q-Q plot shows us how the data is almost next to the line, although some outliers separated from the data are visible. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data fits appropriately to the distribution and the last graph determines that the data set are distributed or exhibit a gamma behavior although some discrepancy between the red line with the foreground in 0.3 approximately. The presence of the table below shows the numerical values of the different goodness of fit tests: logL that presents us with a high value of 695.58. The AIC has a minimum value of -1387.17. BIC also has a low value of -1381.75. The smaller chi square value of 12.6 indicates that it is credible. Also the lower AD value of 0.53 and the KS value of 0.06. All applied GOF-tests confirm that the distributions of the $JS(p,q)$. confirm the gamma probability distribution.

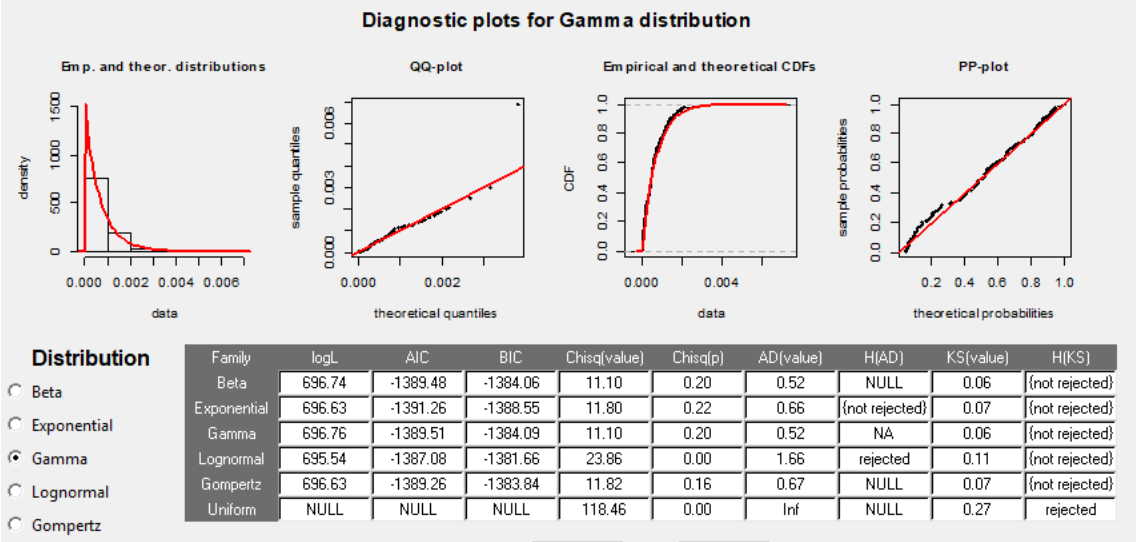
JS(q,p)**:** GOF-test of fit for the Gamma distribution

$$V_n^* = 3.0817,\ Vn = 1.322056,\ p-value = 0.02933$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is not rejected with a level of significance of 0.01.

The fitted gamma distribution has the following parameters:

$$shape : 0.9438537 \qquad rate : 1364.1379998$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	696.74	-1389.48	-1384.06	11.10	0.20	0.52	NULL	0.06	{not rejected}
Exponential	696.63	-1391.26	-1388.55	11.80	0.22	0.66	{not rejected}	0.07	{not rejected}
Gamma	696.76	-1389.51	-1384.09	11.10	0.20	0.52	NA	0.06	{not rejected}
Lognormal	695.54	-1387.08	-1381.66	23.86	0.00	1.66	rejected	0.11	{not rejected}
Gompertz	696.63	-1389.26	-1383.84	11.82	0.16	0.67	NULL	0.07	{not rejected}
Uniform	NULL	NULL	NULL	118.46	0.00	Inf	NULL	0.27	rejected

Fig. 2.11 gamma distribution: $JS(q,p)$

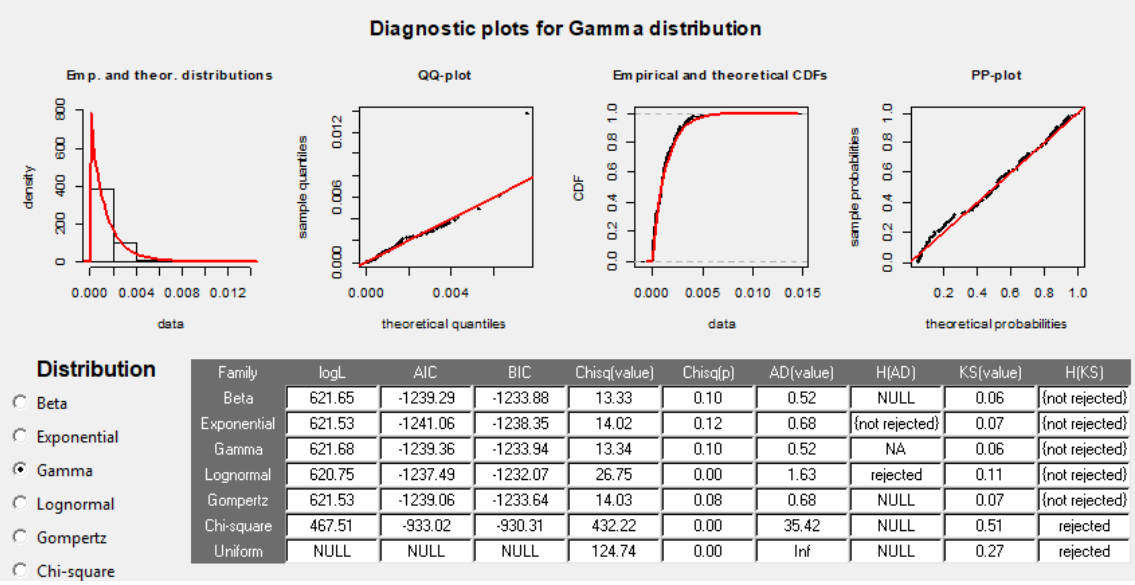
We see that in the Figure 2.11 the null hypothesis is not rejected, the data show a behavior approximately gamma for the similarity function of Jensen-Shannon $JS(q,p)$ where the first graph shows the density of the gamma distribution. On the second graph, the Q-Q plot visually determines that the data is almost close to the line, although outliers are displayed separately from the data. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data fits appropriately to the distribution and the last graph determines that the data set are distributed or exhibit a gamma behavior although some discrepancy between the red line with the foreground in 0.3 approximately. Also, it presents a table in the lower part where the numerical values of the different goodness of fit tests are observed: logL has a high value of 696.76. The AIC has a minimum value of -1389.51. BIC also has a low value of -1384.09. The smaller chi square value of 11.1 tells us that it fits the distribution. Also the lower AD value of 0.52 and the KS value of 0.06. All used GOF-tests confirm that the distributions of the quasi-distances is approximated well by gamma distribution.

c) Skew divergence (Sα):
S$\alpha(p,q)$: GOF-test of fit for the Gamma distribution

$$V_n^* = 3.3025,\ Vn = 1.345953,\ p - value = 0.01953$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is not rejected with a level of significance of 0.01.
The fitted gamma distribution has the following parameters:

$$shape = 0.9383241 \qquad rate = 689.1993274$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	621.65	-1239.29	-1233.88	13.33	0.10	0.52	NULL	0.06	{not rejected}
Exponential	621.53	-1241.06	-1238.35	14.02	0.12	0.68	{not rejected}	0.07	{not rejected}
Gamma	621.68	-1239.36	-1233.94	13.34	0.10	0.52	NA	0.06	{not rejected}
Lognormal	620.75	-1237.49	-1232.07	26.75	0.00	1.63	rejected	0.11	{not rejected}
Gompertz	621.53	-1239.06	-1233.64	14.03	0.08	0.68	NULL	0.07	{not rejected}
Chi-square	467.51	-933.02	-930.31	432.22	0.00	35.42	NULL	0.51	rejected
Uniform	NULL	NULL	NULL	124.74	0.00	Inf	NULL	0.27	rejected

Fig. 2.12 gamma distribution: $S\alpha(p,q)$

The Figure 2.12 shows the analysis of the similarity function of Skew divergence $S_\alpha(p,q)$ here the first graph shows the density of the gamma distribution. On the second graph Q-Q plot it is visualized that the data is almost next to the straight line although there is outliers in the data. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data fits appropriately to the distribution and the last graph determines that the data set are distributed or exhibit a gamma behavior although some discrepancy between the red line in the foreground at approximately 0.3. A table at the bottom displays the numerical values of the different goodness of fit tests are observed: log likelihood (logL) has a high value of 621.68. The AIC has a minimum value of -1239.36. BIC also has a low value of -1233.94. The smaller chi-square value of 13.34 tells us that it fits the distribution. Also the lower AD value of 0.52 and the KS value of 0.06. All the GOF-tests confirm the gamma

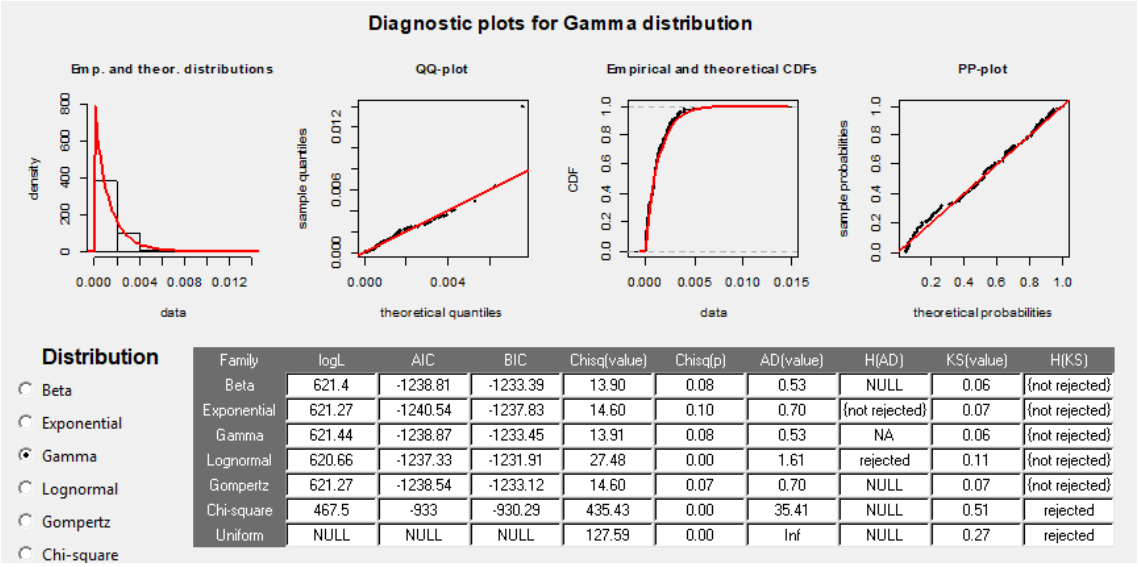
distribution fit for empirical Skew divergence $S_\alpha(p,q)$.

$S\alpha(q,p)$**:** GOF-test for the Gamma distribution

$$V_n^* = 3.4259,\ Vn = 1.359431,\ p-value = 0.01541$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is not rejected with a level of significance of 0.01.
The fitted gamma distribution has the following parameters:

$$shape = 0.9353938 \qquad rate = 685.6796419$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	621.4	-1238.81	-1233.39	13.90	0.08	0.53	NULL	0.06	{not rejected}
Exponential	621.27	-1240.54	-1237.83	14.60	0.10	0.70	{not rejected}	0.07	{not rejected}
Gamma	621.44	-1238.87	-1233.45	13.91	0.08	0.53	NA	0.06	{not rejected}
Lognormal	620.66	-1237.33	-1231.91	27.48	0.00	1.61	rejected	0.11	{not rejected}
Gompertz	621.27	-1238.54	-1233.12	14.60	0.07	0.70	NULL	0.07	{not rejected}
Chi-square	467.5	-933	-930.29	435.43	0.00	35.41	NULL	0.51	rejected
Uniform	NULL	NULL	NULL	127.59	0.00	Inf	NULL	0.27	rejected

Fig. 2.13 gamma distribution: $S\alpha(q,p)$

Figure 2.13 describes the analysis of the similarity function of Skew divergence $S_\alpha(q,p)$ the first graph displays the density of the gamma distribution. On the second graph Q-Q plot allows us to observe that the data is almost next to the straight line although there is outliers in the data. On the third graph we perceive that the dotted line closely follows the adjusted distribution line so that the data fit appropriately to the distribution and the last graph shows that the data show a gamma behavior although some discrepancy is observed between the red line and the first plane at about 0.3. A table displays the numerical values of the different goodness of fit tests are observed: logL has a high value of 621.44. The AIC has a minimum value of -1238.87. BIC

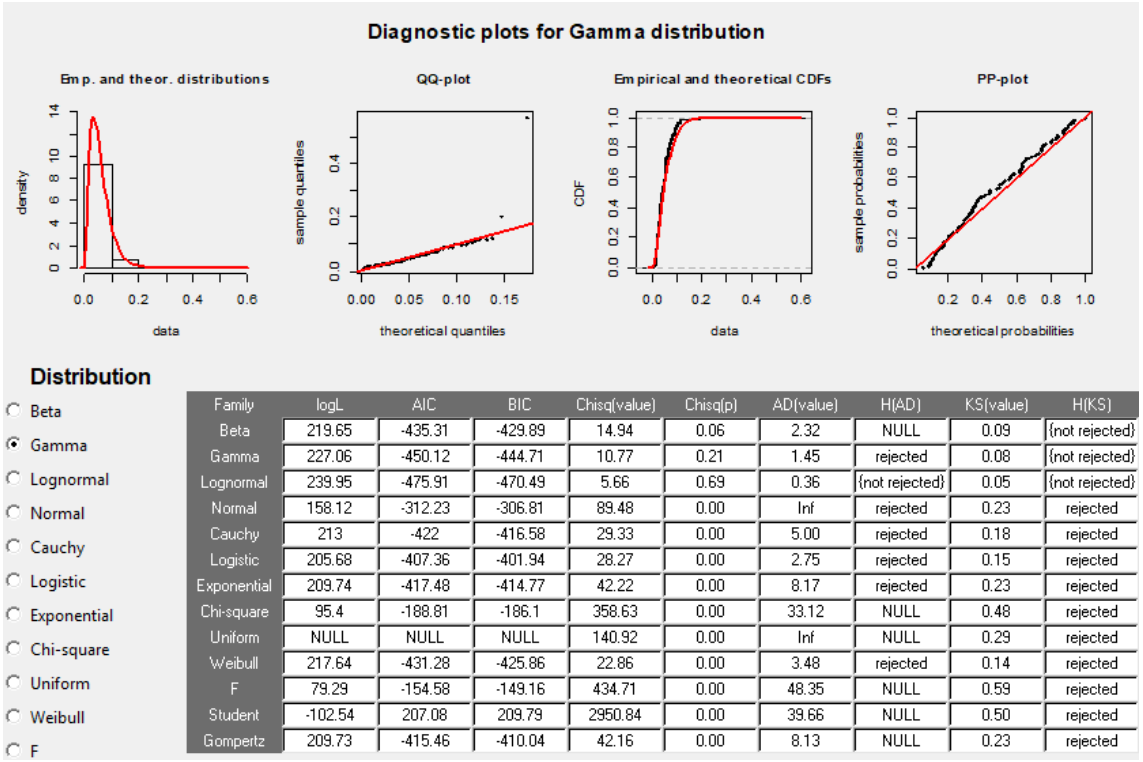
also has a low value of -1233.45. The lower chi square value of 13.91 indicates that it fits the distribution. Also the lower AD value of 0.53 and the KS value of 0.06. All GOF-tests confirm gamma distribution of Skew divergence $S_\alpha(q,p)$.

d) Euclidean $(euc(p,q))$:
$euc(p,q)$**:** GOF-test for the Gamma distribution

$$V_n^* = 16.051,\ Vn = 2.123836,\ p-value < 2.2e-16$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is rejected with a level of significance of 0.01.
Fitted gamma distribution has the following parameters:

$$shape = 2.280293 \qquad rate = 41.011414$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	219.65	-435.31	-429.89	14.94	0.06	2.32	NULL	0.09	{not rejected}
Gamma	227.06	-450.12	-444.71	10.77	0.21	1.45	rejected	0.08	{not rejected}
Lognormal	239.95	-475.91	-470.49	5.66	0.69	0.36	{not rejected}	0.05	{not rejected}
Normal	158.12	-312.23	-306.81	89.48	0.00	Inf	rejected	0.23	rejected
Cauchy	213	-422	-416.58	29.33	0.00	5.00	rejected	0.18	rejected
Logistic	205.68	-407.36	-401.94	28.27	0.00	2.75	rejected	0.15	rejected
Exponential	209.74	-417.48	-414.77	42.22	0.00	8.17	rejected	0.23	rejected
Chi-square	95.4	-188.81	-186.1	358.63	0.00	33.12	NULL	0.48	rejected
Uniform	NULL	NULL	NULL	140.92	0.00	Inf	NULL	0.29	rejected
Weibull	217.64	-431.28	-425.86	22.86	0.00	3.48	rejected	0.14	rejected
F	79.29	-154.58	-149.16	434.71	0.00	48.35	NULL	0.59	rejected
Student	-102.54	207.08	209.79	2950.84	0.00	39.66	NULL	0.50	rejected
Gompertz	209.73	-415.46	-410.04	42.16	0.00	8.13	NULL	0.23	rejected

Fig. 2.14 gamma distribution: $euc(p,q)$

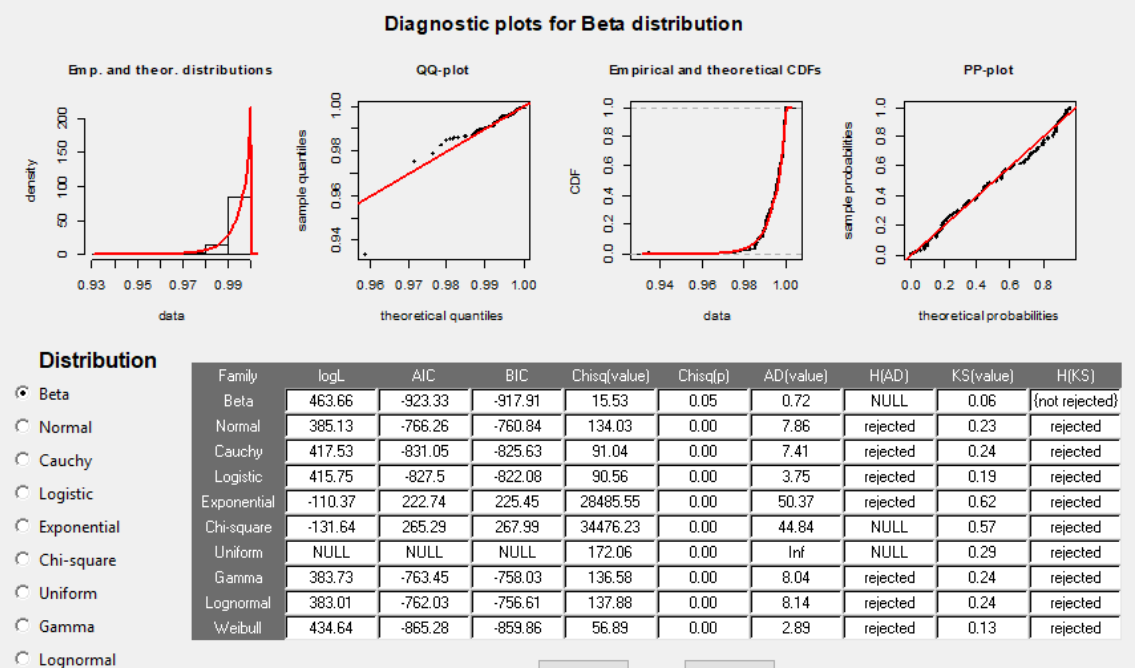
Figure 2.14 shows the analysis of the similarity function of the Euclidean $euc(p,q)$ where the first graph shows the density of the gamma distribution. On the second graph Q-Q plot visually determines that the data is closer to the straight line although there are outliers in the data. On the third graph we perceive that the dotted line closely follows the adjusted distribution line, so that the data fits appropriately to the distribution and the last graph shows that the data show a gamma behavior although some discrepancies are observed between the red line and the first plane at approximately 0.5 and 0.7. In turn, it presents a table in the lower part where the numerical values of the different goodness of fit tests are observed: logL has a high value of 227.06. The AIC has a minimum value of -450.12. BIC also has a low value of -444.71. The smaller chi-square value of 10.77 tells us that it fits the distribution. Also the lower AD value of 1.45 and the KS value of 0.08. All GOF-tests confirms the gamma distribution of $euc(p,q)$.

e) $cos(p,q)$**:** GOF-test for the Gamma distribution

$$V_n^* = -25.823,\ Vn = 0.990091,\ p-value < 2.2e-16$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is rejected with a level of significance of 0.01.
The fitted Beta distribution has the following parameters:

$$shape1 = 151.4651969 \qquad shape2 = 0.8655767$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	463.66	-923.33	-917.91	15.53	0.05	0.72	NULL	0.06	{not rejected}
Normal	385.13	-766.26	-760.84	134.03	0.00	7.86	rejected	0.23	rejected
Cauchy	417.53	-831.05	-825.63	91.04	0.00	7.41	rejected	0.24	rejected
Logistic	415.75	-827.5	-822.08	90.56	0.00	3.75	rejected	0.19	rejected
Exponential	-110.37	222.74	225.45	28485.55	0.00	50.37	rejected	0.62	rejected
Chi-square	-131.64	265.29	267.99	34476.23	0.00	44.84	NULL	0.57	rejected
Uniform	NULL	NULL	NULL	172.06	0.00	Inf	NULL	0.29	rejected
Gamma	383.73	-763.45	-758.03	136.58	0.00	8.04	rejected	0.24	rejected
Lognormal	383.01	-762.03	-756.61	137.88	0.00	8.14	rejected	0.24	rejected
Weibull	434.64	-865.28	-859.86	56.89	0.00	2.89	rejected	0.13	rejected

Fig. 2.15 beta distribution: $cos(p,q)$

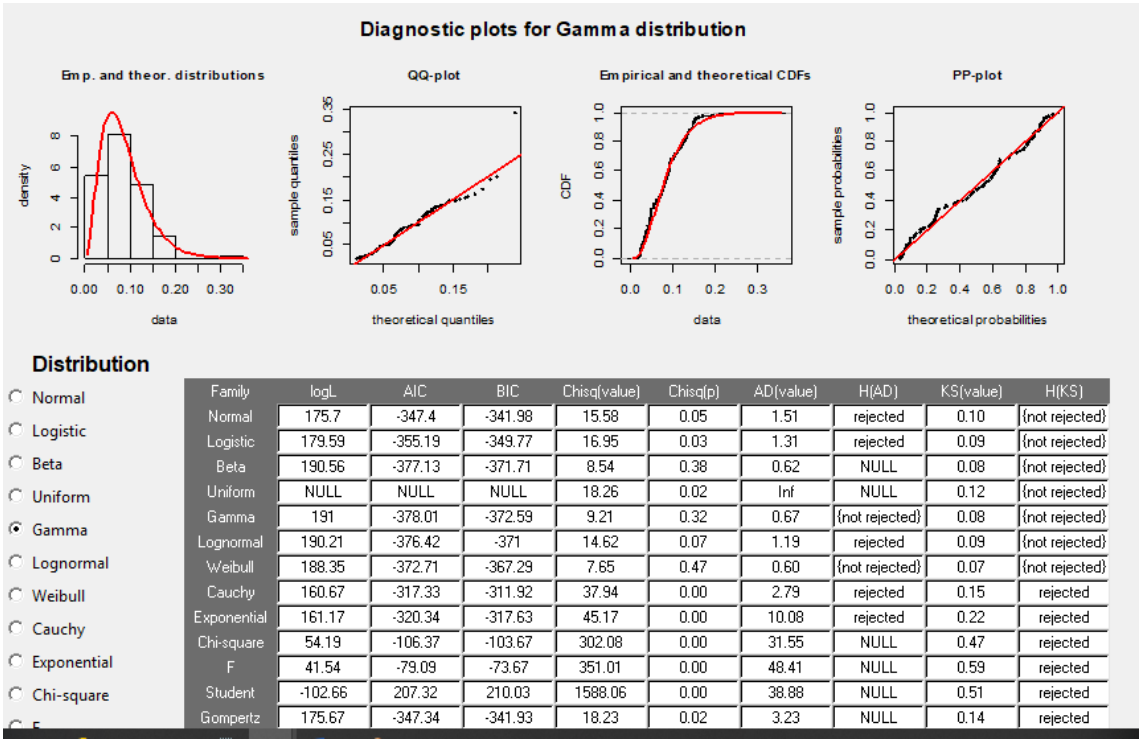
Figure 2.15 shows the analysis of the cosine similarity function $cos(p,q)$ here the first graph shows the density of the Beta distribution. On the second graph Q-Q plot it is visualized that the data are almost next to the straight line although there are some points that are scattered. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data adequately fits the distribution and the last graph determines that the data set are distributed or have a Beta behavior although some discrepancy between the red line and the foreground in approximately from 0.6 to 0.8. A table displays the numerical values of the different goodness of fit tests are observed: logL has a high value of 463.66. The AIC has a minimum value of -923.33. BIC also has a low value of -917.91. The smaller chi-square value of 15.53 tells us that it fits the distribution. Also the lower AD value of 0.72 and the KS value of 0.06. All GOF-tests confirms the Beta distribution for cosine similarity function $cos(p,q)$.

f) L1 ($L_1(p,q)$):
GOF-test for the Gamma distribution

$$V_n^* = 0.64427,\ Vn = 1.044253,\ p-value = 0.6487$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is not rejected with a level of significance of 0.01.
The fitted gamma distribution has the following parameters:

$$shape = 3.145446 \qquad rate = 36.518563$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	175.7	-347.4	-341.98	15.58	0.05	1.51	rejected	0.10	{not rejected}
Logistic	179.59	-355.19	-349.77	16.95	0.03	1.31	rejected	0.09	{not rejected}
Beta	190.56	-377.13	-371.71	8.54	0.38	0.62	NULL	0.08	{not rejected}
Uniform	NULL	NULL	NULL	18.26	0.02	Inf	NULL	0.12	{not rejected}
Gamma	191	-378.01	-372.59	9.21	0.32	0.67	{not rejected}	0.08	{not rejected}
Lognormal	190.21	-376.42	-371	14.62	0.07	1.19	rejected	0.09	{not rejected}
Weibull	188.35	-372.71	-367.29	7.65	0.47	0.60	{not rejected}	0.07	{not rejected}
Cauchy	160.67	-317.33	-311.92	37.94	0.00	2.79	rejected	0.15	rejected
Exponential	161.17	-320.34	-317.63	45.17	0.00	10.08	rejected	0.22	rejected
Chi-square	54.19	-106.37	-103.67	302.08	0.00	31.55	NULL	0.47	rejected
F	41.54	-79.09	-73.67	351.01	0.00	48.41	NULL	0.59	rejected
Student	-102.66	207.32	210.03	1588.06	0.00	38.88	NULL	0.51	rejected
Gompertz	175.67	-347.34	-341.93	18.23	0.02	3.23	NULL	0.14	rejected

Fig. 2.16 gamma distribution: $L_1(p,q)$

Figure 2.16 presents the analysis of the similarity function L_1. Where the first graph of $L_1(p,q)$ displays the density of the gamma distribution. On the second graph, the Q-Qplot visually determines that the data is almost close to the line, although outliers are displayed separately from the data. On the third graph it is observed that the dotted line closely follows the adjusted distribution line so that the data fits appropriately to the distribution and the last graph determines that the data set are distributed or exhibit a gamma behavior although some discrepancy between the red line and the foreground at about 0.35. A table in the lower part displays the numerical values of the different goodness of fit tests are observed: logL has a high value of 191. The AIC has a minimum value of -378.01. BIC also has a low value of -372.59. The lower chi square value of 9.21 indicates that it fits the distribution. Also the lower AD

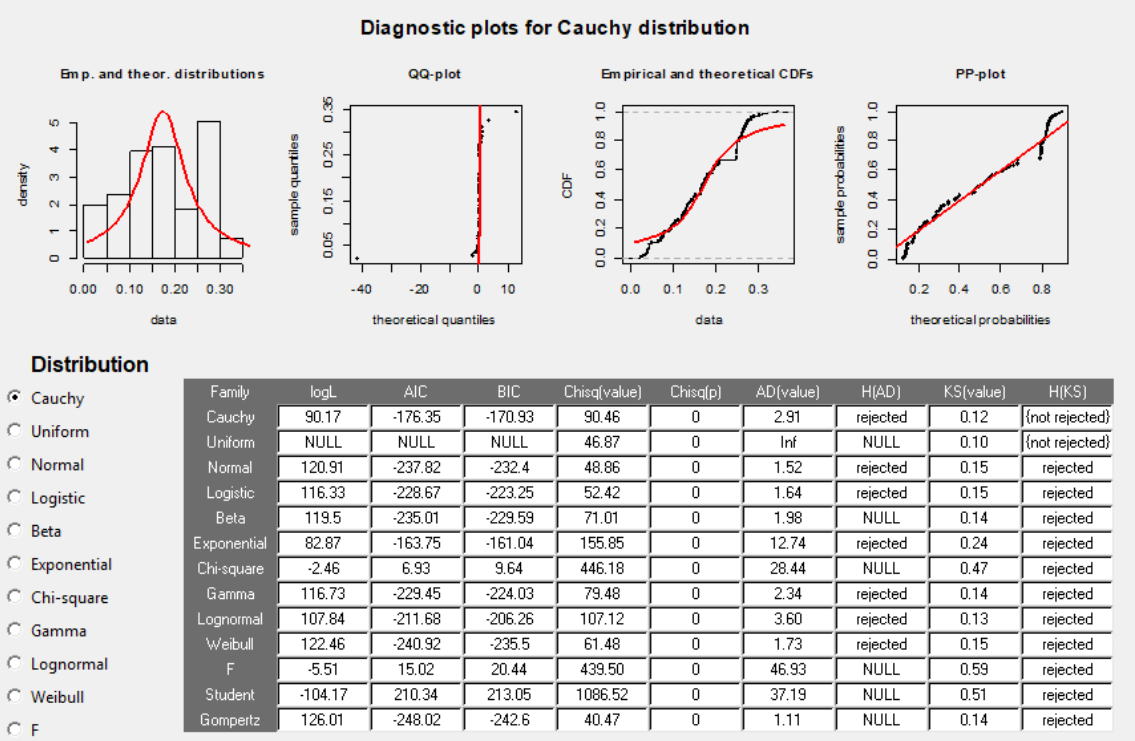
value of 0.67 and the KS value of 0.08. All GOF-tests confirm the gamma probability distribution of the similarity function L_1.

g) $conf(p,q)$**:** GOF-test for the gamma distribution

$$V_n^* = -3.9963,\ Vn = 0.8085334,\ p-value = 0.004716$$

we calculate the decision rule of a goodness of fit test for H_0 based on the ratio of two estimators of the population variance (Vn) indicating that the null hypothesis is rejected with a level of significance of 0.01.
The fitted Cauchy distribution has the following parameters:

$$location = 0.17332712 \qquad scale = 0.05849009$$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Cauchy	90.17	-176.35	-170.93	90.46	0	2.91	rejected	0.12	{not rejected}
Uniform	NULL	NULL	NULL	46.87	0	Inf	NULL	0.10	{not rejected}
Normal	120.91	-237.82	-232.4	48.86	0	1.52	rejected	0.15	rejected
Logistic	116.33	-228.67	-223.25	52.42	0	1.64	rejected	0.15	rejected
Beta	119.5	-235.01	-229.59	71.01	0	1.98	NULL	0.14	rejected
Exponential	82.87	-163.75	-161.04	155.85	0	12.74	rejected	0.24	rejected
Chi-square	-2.46	6.93	9.64	446.18	0	28.44	NULL	0.47	rejected
Gamma	116.73	-229.45	-224.03	79.48	0	2.34	rejected	0.14	rejected
Lognormal	107.84	-211.68	-206.26	107.12	0	3.60	rejected	0.13	rejected
Weibull	122.46	-240.92	-235.5	61.48	0	1.73	rejected	0.15	rejected
F	-5.51	15.02	20.44	439.50	0	46.93	NULL	0.59	rejected
Student	-104.17	210.34	213.05	1086.52	0	37.19	NULL	0.51	rejected
Gompertz	126.01	-248.02	-242.6	40.47	0	1.11	NULL	0.14	rejected

Fig. 2.17 cauchy distribution: $conf(p,q)$

Figure 2.17 shows the analysis of the Confusion similarity function $conf(p,q)$ where the first graph shows the density of the Cauchy distribution. On the second graph Q-Q plot visually determines that the data is closer to the straight line although there are

outliers in the data. On the third graph we perceive that the stepped line follows the adjusted distribution line so the data adjust to the distribution and the last graph shows that the data present a Cauchy behavior although there are some strong discrepancies between the red line and the first one. A table displays the numerical values of the different goodness of fit tests are observed: log likelihood (logL) has a high value of 90.17. The AIC has a minimum value of -176.35. BIC also has a low value of -170.93. The smaller chi-squared value of 90.46 indicates that it fits the distribution. Also the lower AD value of 2.91 and the KS value of 0.12. All GOF-tests confirm the Cauchy distribution for Confusion similarity function $conf(p,q)$.

(Wasserstein et al., 2019) consider that the results should not be categorized based on an arbitrary p- value. They recommend not using the phrase "statistically significant" since it is illogical and inappropriate to dichotomize the results as "significant" and "not significant" since it is not an arbitrator of the truth. It is better to analyze and report all relevant data and results. In turn, he recommends accepting uncertainty and being reflexive, open and modest. Emphasizing that uncertainty must be accepted and embracing variation in effects. Likewise (Greenland, 2019), recommends us to replace the statistically significant sentence by p-value as an equality. For example $p = 0.027$ since they depend on the size of the sample.

Therefore, having obtained different results in the analysis of the test of the null hypothesis with the measure of variance ratio (where the p value of one is above 0.01 and the other below this value, which would indicate an not reject and the other a reject of the null hypothesis) and the others GOF-test. This discordance of results can be seen answered and explained in Wasserstein et al. (2019) paper, since this study explains that the arbitrary threshold p is not an absolute to decide and/or to indicate if the results are statistically significant, since the nature of the data provide noisy signals and that variation is the cause of uncertainty. Therefore, it is better to name the p value without highlighting or using the words statistically significant and presenting all the relevant findings found. It is like that, all these scientific arguments make us reflect that p values are not an arbitrator of absolute truth to make final decisions but a probability value.

Chapter 3

Close to uniform distribution

Uniform distribution does not usually occur in nature, but it is a very important reference distribution for the analysis, study and understanding of many real cases due to the simplicity of the model. This distribution can have a continuous and discrete versions that model a range of values with equal probability, that is, they present a constant probability.

A random variable has a uniform distribution when each value of the random variable is equally likely, and values are uniformly distributed throughout some interval. In the next section we consider the uniform distributions discrete and continuous.

3.1 Uniform distribution

3.1.1 Discrete uniform distribution

A random variable X follows the discrete uniform distribution on the interval $[a, a+1, ..., b]$, if it may attain each of these values with equal probability. We have then:
Density function:

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{for} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Distribution function:

$$F_X(x) = P(X \leq x) \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a+1}{n} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > b \end{cases} \tag{3.2}$$

3.1.2 Continuous uniform distribution

The continuous uniform distribution on the interval is one of the simplest of all probability distributions, but nonetheless very important. This distribution is the basic tool for simulating other probability distributions. The uniform distribution corresponds to picking a point at random from the interval. Also this distribution is known as a rectangular distribution and that has constant probability. The distribution is often abbreviated like $U(a,b)$
Density function:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

Distribution function:

$$F_X(x) = P(X \leq x) \begin{cases} 0 & \text{if} \quad x < a \\ \frac{x-a}{b-a} & \text{if} \quad a \leq x < b \\ 1 & \text{if} \quad x \geq b \end{cases} \tag{3.4}$$

3.1.3 Results of GOF test for uniform distribution for quasi-distances between two corpuses

Experiment 2: Lowest frequency of occurrence of words

Here we analyzed 35 key words with the lowest frequency in the historical corpus and by uniform behavior since this would lead us to the selection of points with small discrepancies between all possible designs.

To start experiment 2, we first obtained the values of the quasi-distances for the keywords. Locating each keyword in its corresponding paragraph to obtain the relative frequency of the word with the words of the paragraph to which it belongs. There were cases that the keyword was repeated twice in the same paragraph and its calculation was in its proportion. Then this numerical value of the sum of all the numerical values would be the relative frequency of each word with the corpus of Lasso and Rojas. These data were calculated for each keyword used in similarity measures. For example the keyword niiyo (my essence) appears only once in paragraphs 37 and 58 each paragraph has 11 and 20 words respectively. Then, we obtain its relative frequency $1/11 = 0.091$ and $1/20 = 0.05$ its sum (0.141) is obtained and then calculate the relative frequency of each value of the word between the sum of the values ($0.091/0.141 = 0.645$, case of paragraph 37 of the Lasso corpus). Finally these are the values that are calculated for the similarity functions (see formulas in Table 2.1) for each

keyword. Experiment 2 analyzed the words with less frequency of occurrence, that is, the keywords that were not included in several paragraphs, these being 35 of the 111 key words.

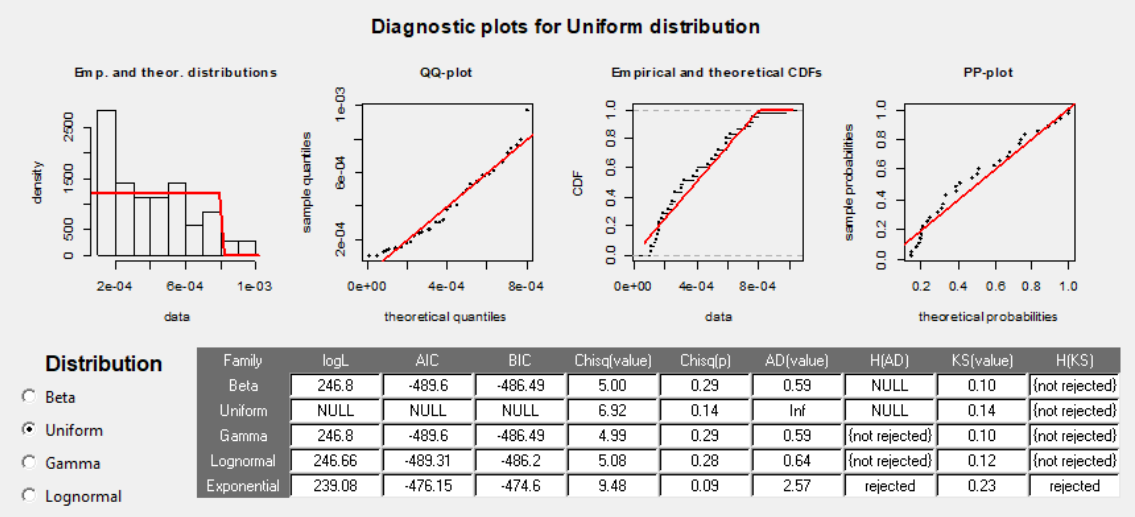
In Table 3.1, we show results of different similarity functions for uniform probability distributions, where the second column (Parameter) is subdivided into min (minimum value of the data range) and max (maximum value of the data range). The third column (Fitted parameter) subdivided into min * (minimum) and max * (maximum) which are the adjusted parameters of the uniform distribution and the fourth column labeled KS (Kolmogorov-Smirnov) is subdivided into D (statistic with the rank parameter) and D * (statistical of the adjusted parameters) which is the statistis of this test and is the absolute max distance (supremum) between the CDFs of the two corpus. The closer this number is to 0 the more likely that the two samples were drawn from the same distribution. Here $D = sup_{x \in R} |F_n(x) - F(x)|$

Table 3.1 **Experiment 2**:Comparison of values functions for probability distributions

Functions	Parameter		Fitted parameter		KS	
	min	max	min*	max*	D	D*
KL(p,q)	0.0001056	0.0009767	-8.0149e-06	0.0008027	0.2693	0.1401
KL(q,p)	0.0001060	0.0009900	-8.5496e-06	8.0382e-04	0.2706	0.141
JS(p,q)	2.649e-05	0.0002474	-2.1214e-0.6	2.0090e-04	0.2706	0.491
JS(q,p)	2.639e-05	0.0002441	-1.987e-06	2.0062e-04	0.27	0.1401
$S_\alpha(p,q)$	5.18e-05	0.0004810	-3.9705e-06	3.9339e-0.4	0.2698	0.496
$S_\alpha(q,p)$	5.1836e-05	0.0004823	-4.0246e-06	3.9350e-04	0.270	0.141
euc(p,q)	0.0113834	0.04012	0.0115289	0.036572	0.1462	0.113
cos(p,q)	0.9982424	0.9998193	0.9984464	1.0000279	0.2811	0.150
$L_1(p,q)$	0.020154	0.062767	0.0178674	0.05735	0.1892	0.0945
Conf(p,q)	0.045844	0.347059	0.094743	0.32933	0.2061	0.1476

Then, in more detail the figures from 3.1 to 3.20 with their respective interpretations show us the plot of the uniform distribution and cumulative distribution functions for each of the similarities indicated in Table 3.1

a) KL divergence: $D(p||q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	246.8	-489.6	-486.49	5.00	0.29	0.59	NULL	0.10	{not rejected}
Uniform	NULL	NULL	NULL	6.92	0.14	Inf	NULL	0.14	{not rejected}
Gamma	246.8	-489.6	-486.49	4.99	0.29	0.59	{not rejected}	0.10	{not rejected}
Lognormal	246.66	-489.31	-486.2	5.08	0.28	0.64	{not rejected}	0.12	{not rejected}
Exponential	239.08	-476.15	-474.6	9.48	0.09	2.57	rejected	0.23	rejected

Fig. 3.1 uniform distribution: $D(p||q)$

Figure 3.1 includes the analysis of the similarity function of the Kullback leibler divergence $D(p||q)$, where the first graph represents the density of the uniform distribution. On the second graph the Q-Q plot visually determines that most of the data is close to a straight line, although it has the presence of outliers. On the third graph it is observed that the stepped line follows the adjusted distribution line so the data adjust to the distribution and the last graph determines that the data sets are evenly distributed, although some discrepancies are observed between the red line with the first plane between 0.4 and 0.6 approximately. Also below it presents a table that reflects the numerical values of some of the different goodness of fit tests such as: The lower chi square value of 6.92 that tells us that it fits and where the value of the KS discrepancy is 0.14. The values of GOF-tests assert that the distribution of the $D(p||q)$ fits the uniform probabilistic model.

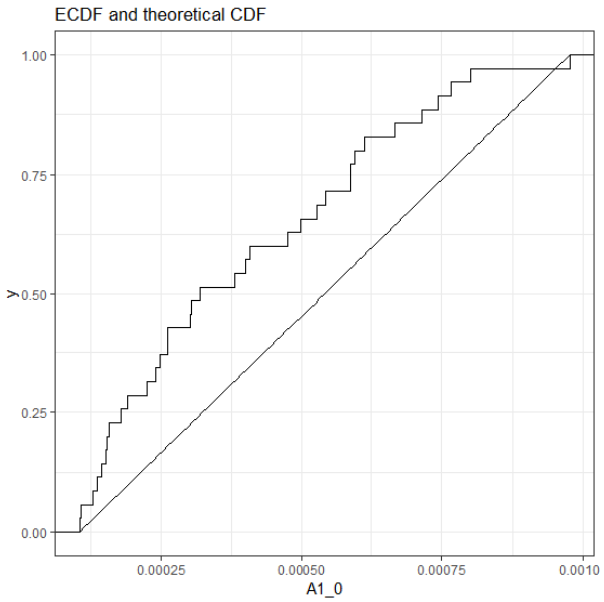


Fig. 3.2 Empirical and theoretical CDFs: $D(p||q)$

Figure 3.2 shows that the staggered line is not close to the adjusted distribution line, so we reject the uniform distribution.

b) KL divergence: $D(q||p)$

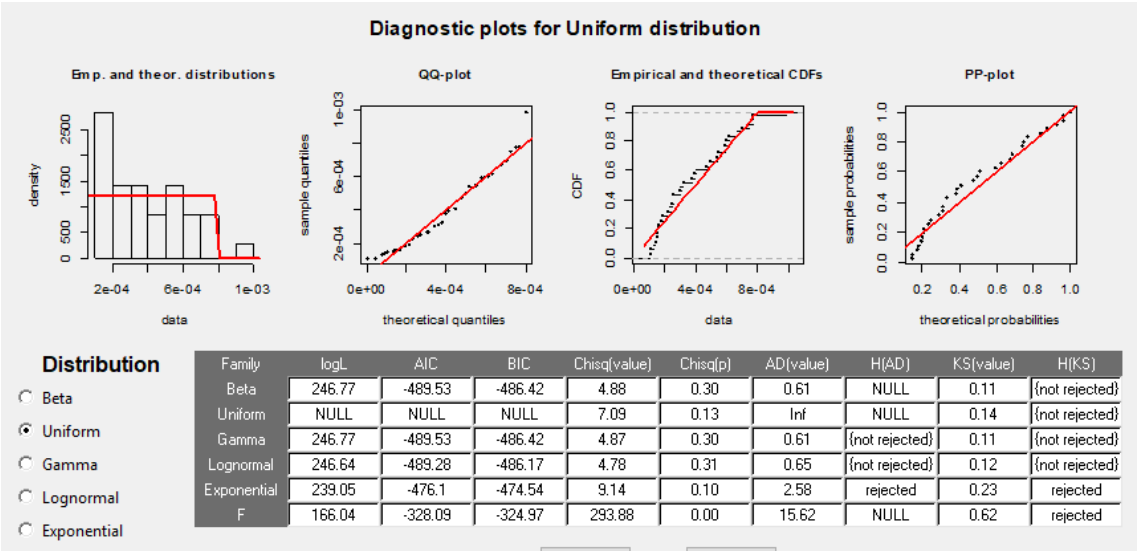


Fig. 3.3 uniform distribution: $D(q||p)$

Figure 3.3 shows the analyzed results of the similarity function of the Kullback leibler divergence $D(q||p)$, the first graph represents the density of the uniform distribution.

On the second graph Q-Q plot reveals that most of the data approaches a straight line although it presents outliers. On the third graph it is observed that the stepped line follows the adjusted distribution line presenting an adequate adjustment and the last graph determines that the data sets are evenly distributed, although some discrepancies are observed between the red line with the foreground between 0.35 and 0.6 approximately. In turn, below it presents a table that reflects the numerical values of some of the different goodness of fit tests such as: The lower chi square value of 7.09 that tells us that it fits and the KS value of 0.14. Consequently, the values of the GOF-tests assert that the distribution of the data conform with the uniform probabilistic model.

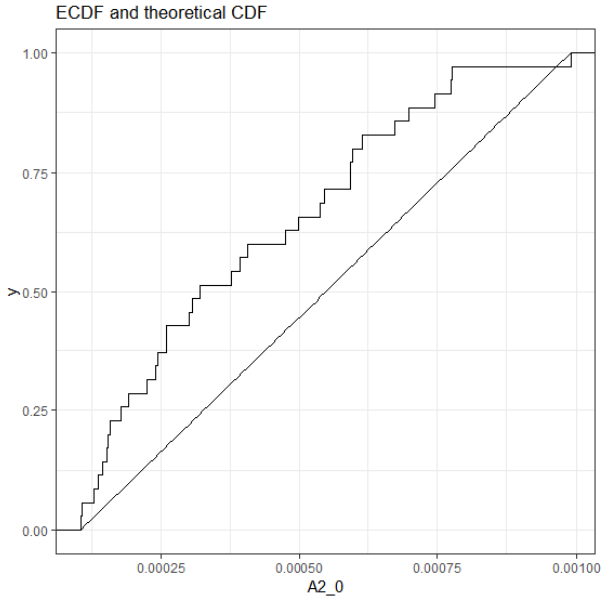
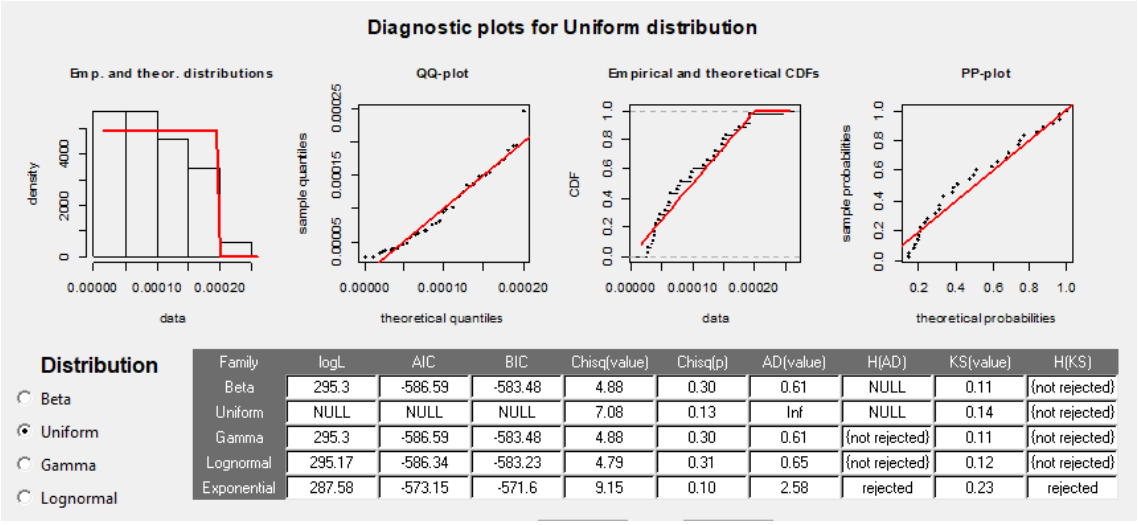


Fig. 3.4 Empirical and theoretical CDFs: $D(q||p)$

Figure 3.4 shows that the staggered line is not close to the adjusted distribution line, so we rejected the uniform distribution.

c) Jensen-Shannon: $JS(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	295.3	-586.59	-583.48	4.88	0.30	0.61	NULL	0.11	{not rejected}
Uniform	NULL	NULL	NULL	7.08	0.13	Inf	NULL	0.14	{not rejected}
Gamma	295.3	-586.59	-583.48	4.88	0.30	0.61	{not rejected}	0.11	{not rejected}
Lognormal	295.17	-586.34	-583.23	4.79	0.31	0.65	{not rejected}	0.12	{not rejected}
Exponential	287.58	-573.15	-571.6	9.15	0.10	2.58	rejected	0.23	rejected

Fig. 3.5 uniform distribution: $JS(p,q)$

Figure 3.5 shows the analysis of the Jensen-Shannon similarity function $JS(p,q)$ exhibits a first graph that shows the density of the uniform distribution. On the second graph Q-Qplot represents how the data is almost close to the line even though outliers are displayed. On the third graph it is observed that the staggered line follows the adjusted distribution line, so the data fits properly to the distribution and the last graph determines that the data set is evenly distributed although some discrepancy is observed between the red line with the first plane between 0.3 and 0.6 approximately. Also, the presence of the table below shows the numerical values of some of the different goodness of fit tests as: the lower chi-square value of 7.08 indicates that it is credible and the KS value of 0.14. All of the GOF-test values determine that the distributions of the data conform with the uniform probability distribution.

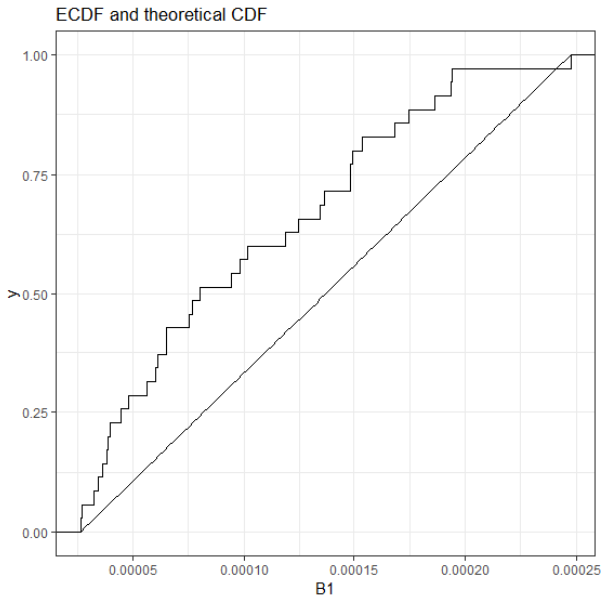


Fig. 3.6 Empirical and theoretical CDFs: $JS(p,q)$

Figure 3.6, shows that the stepped line is on the adjusted distribution line so we rejected the uniform distribution.

d) Jensen-Shannon:$JS(q,p)$

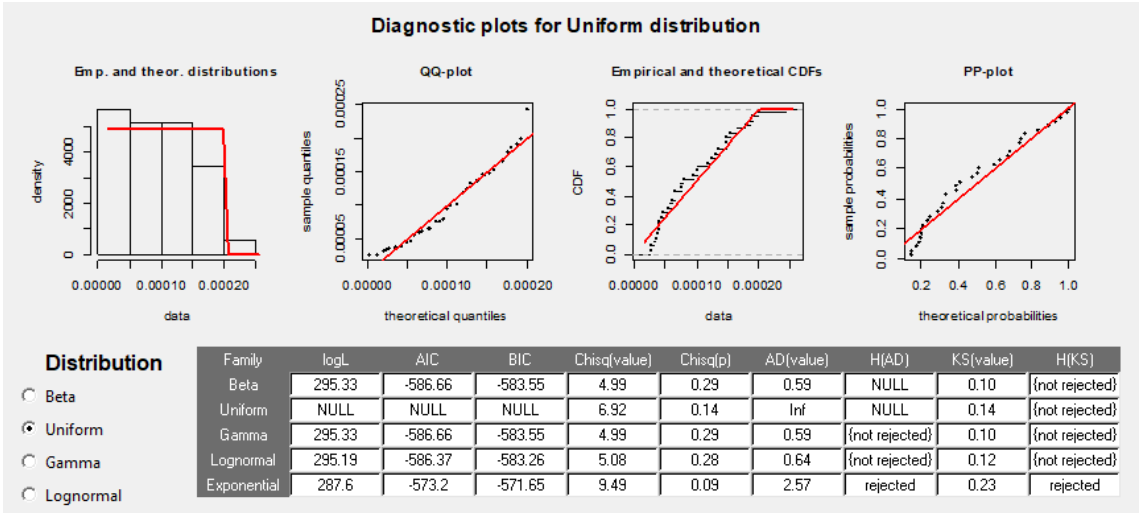


Fig. 3.7 uniform distribution: $JS(q,p)$

Figure 3.7 presents the analysis of the similarity function of Jensen-Shannon $JS(q,p)$ where the first graph shows the density of the uniform distribution. On the second

graph Q-Qplot visually determines that the data is almost close to the line even though it is displayed outliers in the data. On the third graph it is observed that the staggered line closely follows the adjusted distribution line, so the data adequately fits the distribution and the last graph determines that the data set presents a uniform behavior although some discrepancies are observed between the red line with the foreground between 0.35 and 0.6 approximately. Also, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: the lower chi square value of 6.92 indicates that it fits the distribution and the KS value of 0.14. The values of GOF-tests determine that the distribution of the $JS(q,p)$ fits the uniform probability distribution.

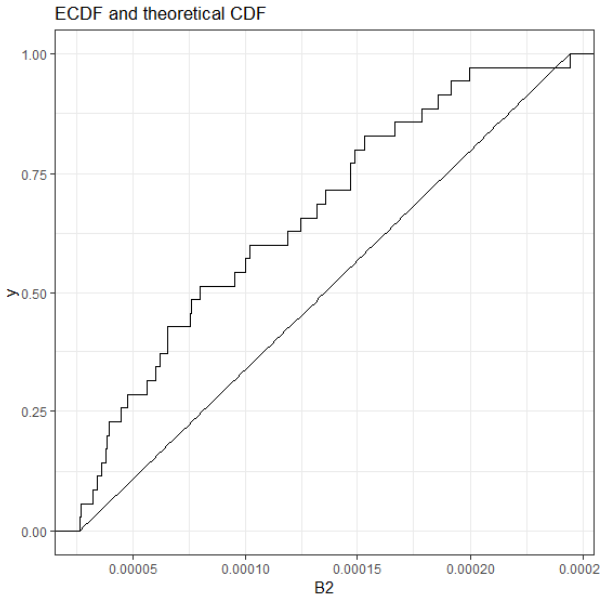
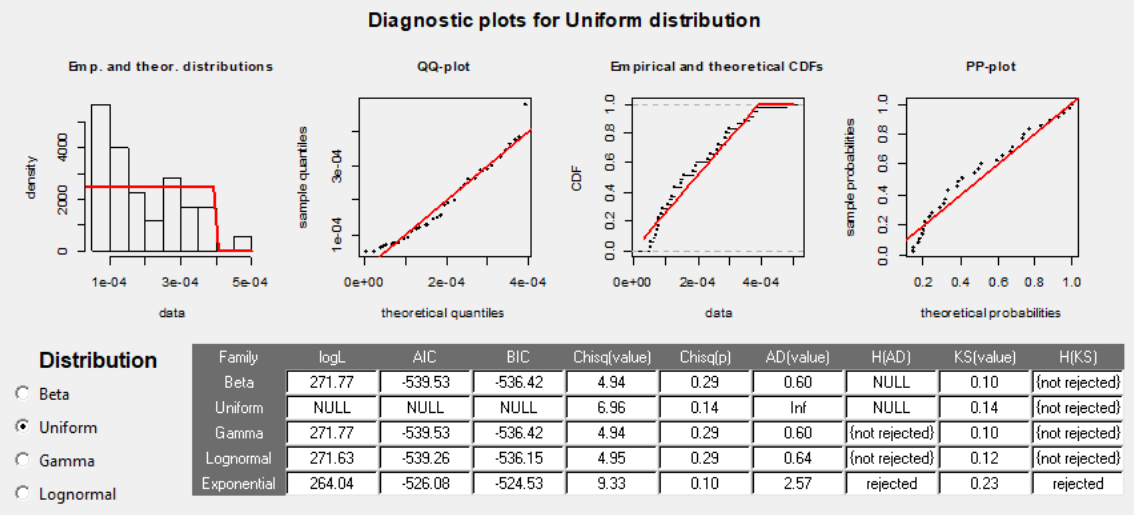


Fig. 3.8 Empirical and theoretical CDFs: $JS(q,p)$

In Figure 3.8, it is observed that the stepped line is on the adjusted distribution line, so we rejected the uniform distribution.

e) Skew divergence:$S_\alpha(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	271.77	-539.53	-536.42	4.94	0.29	0.60	NULL	0.10	{not rejected}
Uniform	NULL	NULL	NULL	6.96	0.14	Inf	NULL	0.14	{not rejected}
Gamma	271.77	-539.53	-536.42	4.94	0.29	0.60	{not rejected}	0.10	{not rejected}
Lognormal	271.63	-539.26	-536.15	4.95	0.29	0.64	{not rejected}	0.12	{not rejected}
Exponential	264.04	-526.08	-524.53	9.33	0.10	2.57	rejected	0.23	rejected

Fig. 3.9 uniform distribution: $S_\alpha(p,q)$

Figure 3.9 shows the analysis of the similarity function of Skew divergence $S_\alpha(p,q)$ here the first graph displays the density of the uniform distribution. On the second graph Q-Q plot it is visualized that the data are almost next to the straight line although there is outliers. On the third graph it is observed that the stepped line follows the adjusted distribution line so that the data fits properly to the distribution and the last graph determines that the data set is uniformly distributed although some discrepancies are observed between the red line and the first plane between 0.37 and 0.6 approximately. In turn, it presents a table in the lower part where the numerical values of some of the different goodness of fit tests are observed as: The smaller chi square value of 6.96 indicates that it fits the distribution and the KS value of 0.14. These values of the adjustment tests determine that the distribution of the $S_\alpha(p,q)$ is adjusted to the uniform probability distribution.

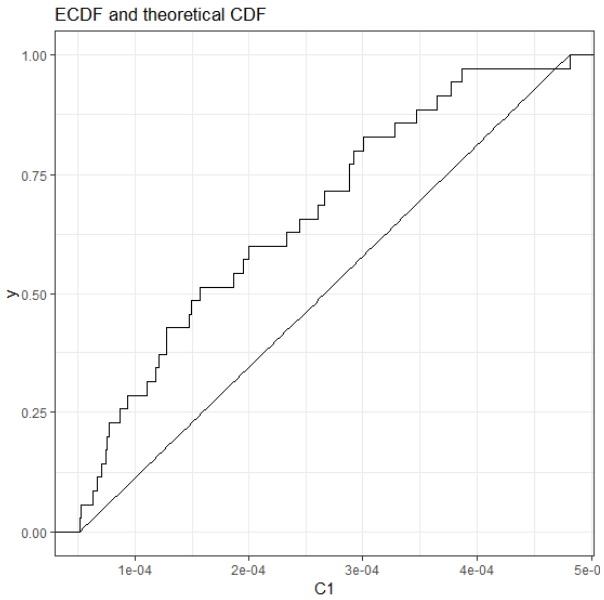


Fig. 3.10 Empirical and theoretical CDFs: $S_\alpha(p,q)$

Figure 3.10 shows a staggered line that is on the adjusted distribution line, so we rejected the uniform distribution.

f) Skew divergence:$S_\alpha(q,p)$

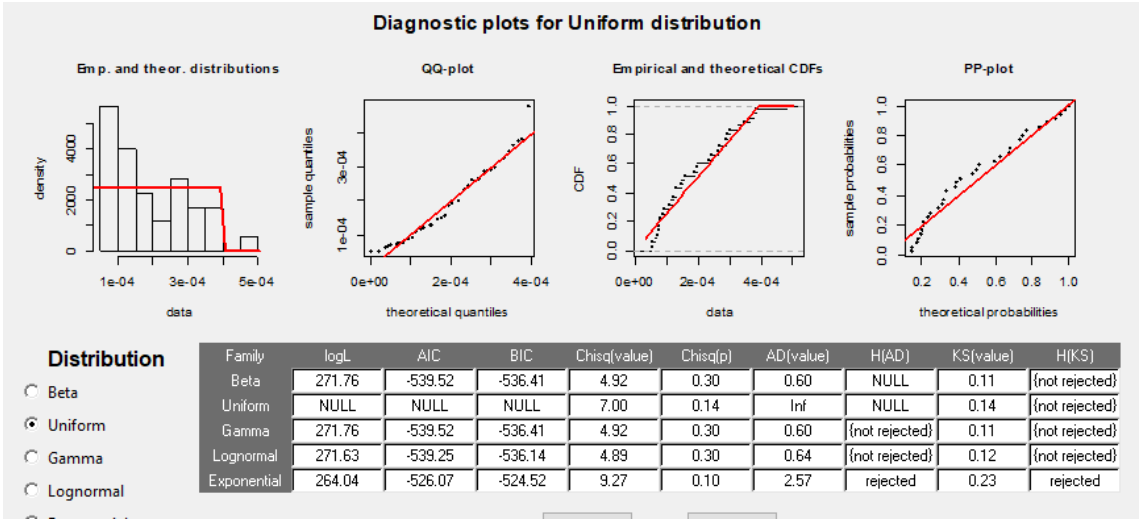


Fig. 3.11 uniform distribution: $S_\alpha(q,p)$

Figure 3.11 describes the analysis of the similarity function of Skew divergence $S_\alpha(q,p)$ the first graph shows the density of the uniform distribution. On the second

graph Q-Q plot allows us to observe that the data is almost close to the straight line although there is outliers. On the third graph we perceive that the staggered line closely follows the adjusted distribution line so that the data fit properly to the distribution and the last graph shows that the data show a uniform behavior although some discrepancies are observed between the red line and the first plane between 0.4 and 0.6 approximately. Also, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: the smaller chi-square value of 7 indicates that it fits the distribution and the KS value of 0.14. These values of GOF-tests allow us to assert that the distributions of the $S_\alpha(q,p)$ fit the uniform probability distribution.

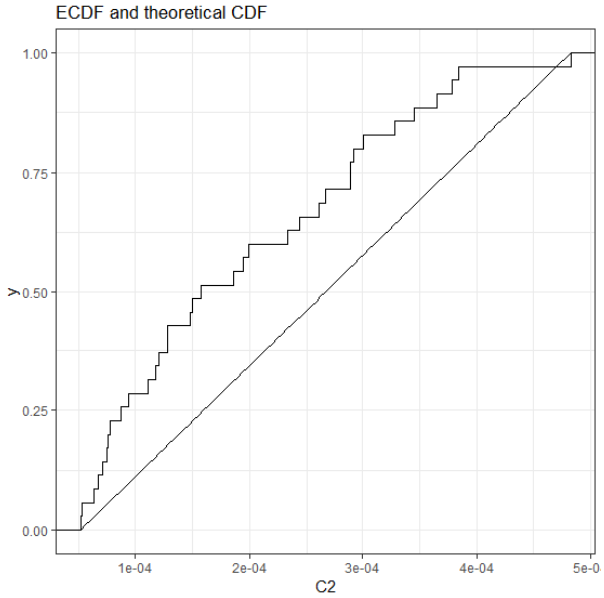
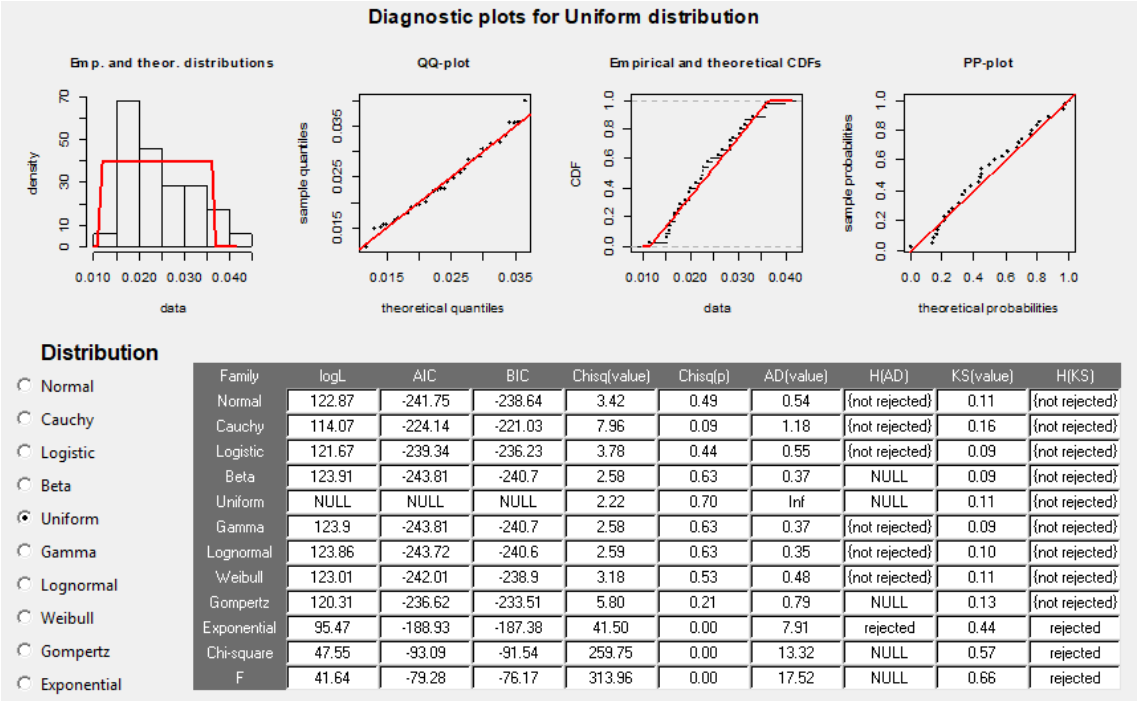


Fig. 3.12 Empirical and theoretical CDFs: $S_\alpha(q,p)$

Figure 3.12 shows how the staggered line is on the adjusted distribution line, so we rejected the uniform distribution.

g) Euclidean:$euc(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	122.87	-241.75	-238.64	3.42	0.49	0.54	{not rejected}	0.11	{not rejected}
Cauchy	114.07	-224.14	-221.03	7.96	0.09	1.18	{not rejected}	0.16	{not rejected}
Logistic	121.67	-239.34	-236.23	3.78	0.44	0.55	{not rejected}	0.09	{not rejected}
Beta	123.91	-243.81	-240.7	2.58	0.63	0.37	NULL	0.09	{not rejected}
Uniform	NULL	NULL	NULL	2.22	0.70	Inf	NULL	0.11	{not rejected}
Gamma	123.9	-243.81	-240.7	2.58	0.63	0.37	{not rejected}	0.09	{not rejected}
Lognormal	123.86	-243.72	-240.6	2.59	0.63	0.35	{not rejected}	0.10	{not rejected}
Weibull	123.01	-242.01	-238.9	3.18	0.53	0.48	{not rejected}	0.11	{not rejected}
Gompertz	120.31	-236.62	-233.51	5.80	0.21	0.79	NULL	0.13	{not rejected}
Exponential	95.47	-188.93	-187.38	41.50	0.00	7.91	rejected	0.44	rejected
Chi-square	47.55	-93.09	-91.54	259.75	0.00	13.32	NULL	0.57	rejected
F	41.64	-79.28	-76.17	313.96	0.00	17.52	NULL	0.66	rejected

Fig. 3.13 uniform distribution: $euc(p,q)$

Figure 3.13 shows the analysis of the similarity function of the Euclidean $euc(p,q)$ where the first graph shows the density of the uniform distribution. On the second graph Q-Q plot visually determines that the data is closer to the straight line although there are outliers. On the third graph we perceive that the staggered line closely follows the adjusted distribution line, so that the data fits appropriately to the distribution and the last graph shows that the data present a uniform behavior, although some discrepancies are observed between the red line and the foreground between 0.4 and 0.6 approximately. In turn, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: The lower chi square value of 2.22 indicates that it fits the distribution and the KS value of 0.11. These values of GOF tests confirm that the $euc(p,q)$ fit the uniform probability distribution.

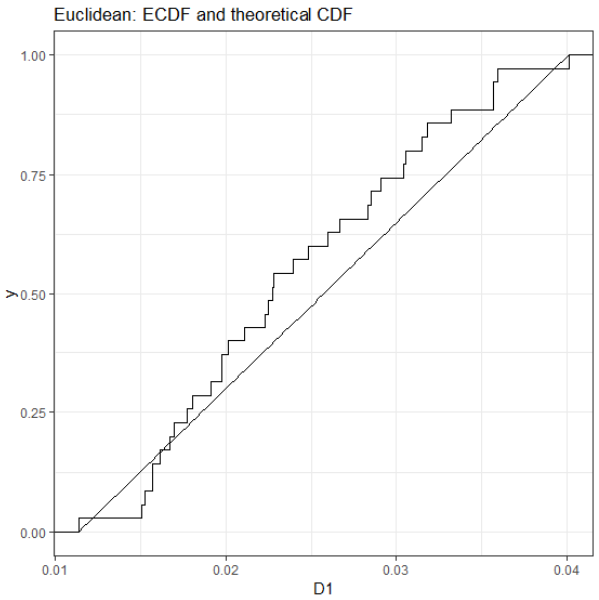
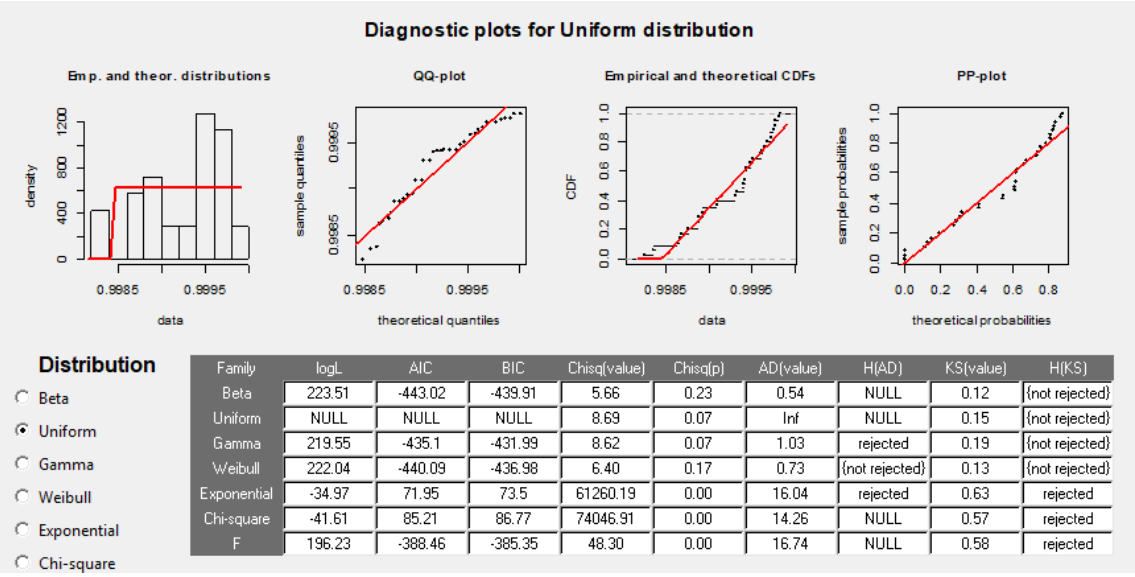


Fig. 3.14 Empirical and theoretical CDFs: $euc(p,q)$

Figure 3.14, allows us to observe that the stepped line follows the adjusted distribution line, so we conclude that the information ranges are adjusted slightly to the uniform distribution.

h) Cosine: $cos(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	223.51	-443.02	-439.91	5.66	0.23	0.54	NULL	0.12	{not rejected}
Uniform	NULL	NULL	NULL	8.69	0.07	Inf	NULL	0.15	{not rejected}
Gamma	219.55	-435.1	-431.99	8.62	0.07	1.03	rejected	0.19	{not rejected}
Weibull	222.04	-440.09	-436.98	6.40	0.17	0.73	{not rejected}	0.13	{not rejected}
Exponential	-34.97	71.95	73.5	61260.19	0.00	16.04	rejected	0.63	rejected
Chi-square	-41.61	85.21	86.77	74046.91	0.00	14.26	NULL	0.57	rejected
F	196.23	-388.46	-385.35	48.30	0.00	16.74	NULL	0.58	rejected

Fig. 3.15 uniform distribution: $cos(p,q)$

Figure 3.15 shows the analysis of the cosine similarity function, $cos(p,q)$ here the first graph shows the density of the uniform distribution. On the second graph Q-Q plot it is visualized that the data are relatively close to the straight line although there are scattered points. On the third graph it is observed that the staggered line follows the adjusted distribution line relatively, so the data is weakly adjusted to the distribution and the last graph determines that the data set is uniformly distributed although some discrepancies are observed between the red line and the foreground between 0.4 to 0.6 approximately. Likewise, it presents a table in the lower part where numerical values of some goodness of fit tests are observed, such as: the lower chi-square value of 8.69 indicates that it fits the distribution and the KS value of 0.15. These values of the adjustment tests determine that the data behave uniformly.

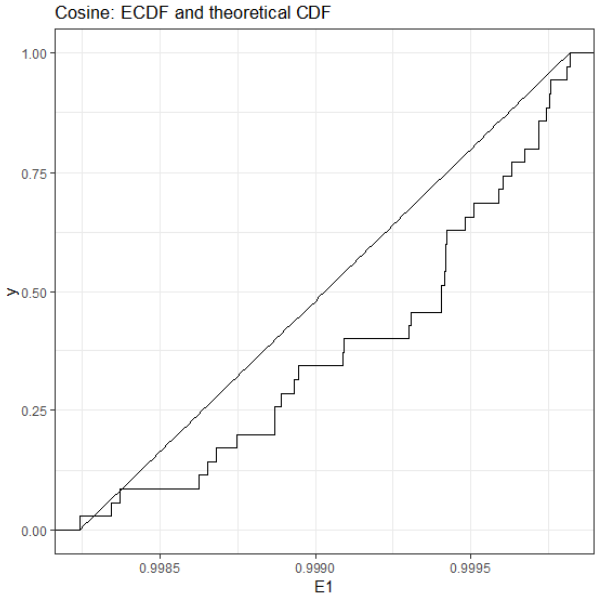
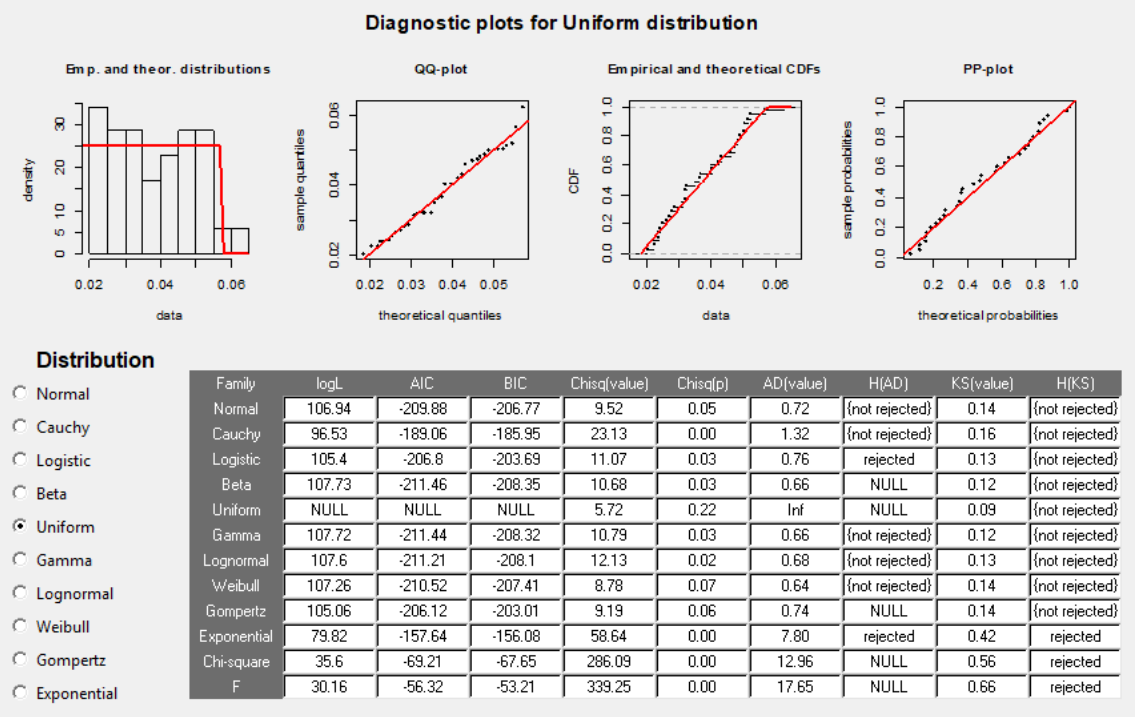


Fig. 3.16 Empirical and theoretical CDFs: $cos(p,q)$

In figure 3.16, it is perceived that the staggered line is below the adjusted distribution line, so we rejected the uniform distribution.

i) L_1: $L_1(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	106.94	-209.88	-206.77	9.52	0.05	0.72	{not rejected}	0.14	{not rejected}
Cauchy	96.53	-189.06	-185.95	23.13	0.00	1.32	{not rejected}	0.16	{not rejected}
Logistic	105.4	-206.8	-203.69	11.07	0.03	0.76	rejected	0.13	{not rejected}
Beta	107.73	-211.46	-208.35	10.68	0.03	0.66	NULL	0.12	{not rejected}
Uniform	NULL	NULL	NULL	5.72	0.22	Inf	NULL	0.09	{not rejected}
Gamma	107.72	-211.44	-208.32	10.79	0.03	0.66	{not rejected}	0.12	{not rejected}
Lognormal	107.6	-211.21	-208.1	12.13	0.02	0.68	{not rejected}	0.13	{not rejected}
Weibull	107.26	-210.52	-207.41	8.78	0.07	0.64	{not rejected}	0.14	{not rejected}
Gompertz	105.06	-206.12	-203.01	9.19	0.06	0.74	NULL	0.14	{not rejected}
Exponential	79.82	-157.64	-156.08	58.64	0.00	7.80	rejected	0.42	rejected
Chi-square	35.6	-69.21	-67.65	286.09	0.00	12.96	NULL	0.56	rejected
F	30.16	-56.32	-53.21	339.25	0.00	17.65	NULL	0.66	rejected

Fig. 3.17 uniform distribution: $L_1(p,q)$

Figure 3.17 presents the analysis of the similarity function L_1. Where the first graph of $L_1(p,q)$ displays the density of the uniform distribution. On the second graph Q-Qplot visually determines that the data is almost close to the line, even though outliers are displayed. On the third graph it is observed that the staggered line closely follows the adjusted distribution line, so the data adequately fits the distribution and the last graph determines that the data set presents a uniform behavior although some discrepancies are observed between the red line and the foreground between 0.4 and 0.6 approximately. Also, it presents a table in the lower part where the numerical values of some goodness of fit tests are observed as: The lower chi square value of 5.72 indicates that it fits the distribution and the KS value of 0.09. All these values of GOF-tests determine that the data fit the uniform probability distribution.

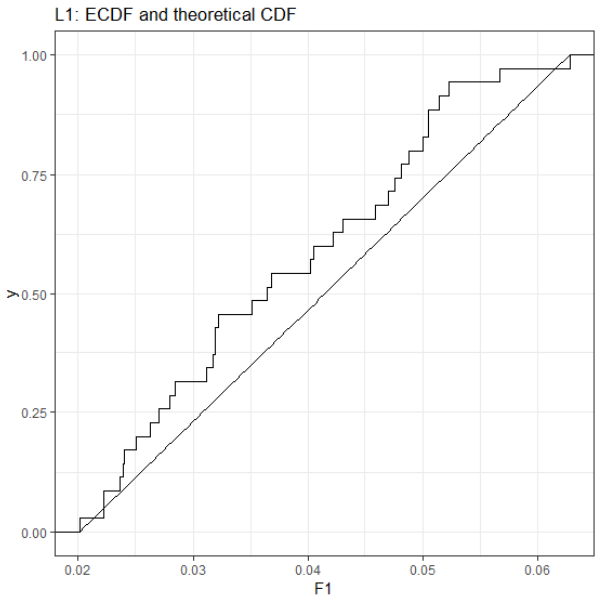
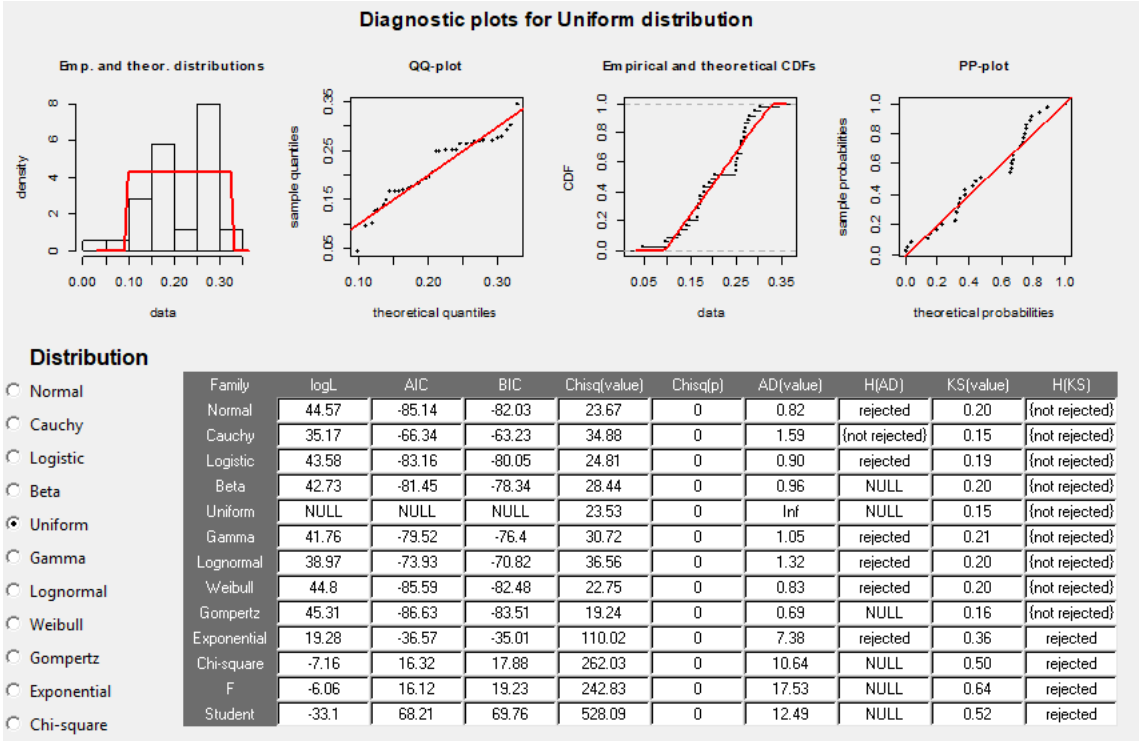


Fig. 3.18 Empirical and theoretical CDFs: $L_1(p,q)$

In figure 3.18, it is visualized that the stepped line is above the adjusted distribution line, so we rejected the uniform distribution.

j) Confusion: $conf(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	44.57	-85.14	-82.03	23.67	0	0.82	rejected	0.20	{not rejected}
Cauchy	35.17	-66.34	-63.23	34.88	0	1.59	{not rejected}	0.15	{not rejected}
Logistic	43.58	-83.16	-80.05	24.81	0	0.90	rejected	0.19	{not rejected}
Beta	42.73	-81.45	-78.34	28.44	0	0.96	NULL	0.20	{not rejected}
Uniform	NULL	NULL	NULL	23.53	0	Inf	NULL	0.15	{not rejected}
Gamma	41.76	-79.52	-76.4	30.72	0	1.05	rejected	0.21	{not rejected}
Lognormal	38.97	-73.93	-70.82	36.56	0	1.32	rejected	0.20	{not rejected}
Weibull	44.8	-85.59	-82.48	22.75	0	0.83	rejected	0.20	{not rejected}
Gompertz	45.31	-86.63	-83.51	19.24	0	0.69	NULL	0.16	{not rejected}
Exponential	19.28	-36.57	-35.01	110.02	0	7.38	rejected	0.36	rejected
Chi-square	-7.16	16.32	17.88	262.03	0	10.64	NULL	0.50	rejected
F	-6.06	16.12	19.23	242.83	0	17.53	NULL	0.64	rejected
Student	-33.1	68.21	69.76	528.09	0	12.49	NULL	0.52	rejected

Fig. 3.19 uniform distribution: $conf(p,q)$

Figure 3.19 shows the analysis of the Confusion similarity function, $conf(p,q)$ where the first graph shows the density of the uniform distribution. On the second graph Q-Q plot visually determines that the data is weakly close to the straight line. On the third graph we perceive that the stepped line follows the adjusted distribution line, so the data is adjusted to the distribution and the last graph shows that the data have a uniform behavior although there are some strong discrepancies between the red line and the first one. flat in 0.5, 0.6 and 0.9 approximately. At the same time, it presents a table in the lower part where the numerical values of some goodness of fit tests are observed as: the lower chi-square value of 23.53 indicates that it fits the distribution and the KS value of 0.15. These values of GOF tests determine that the data behave approximately as uniform distributed.

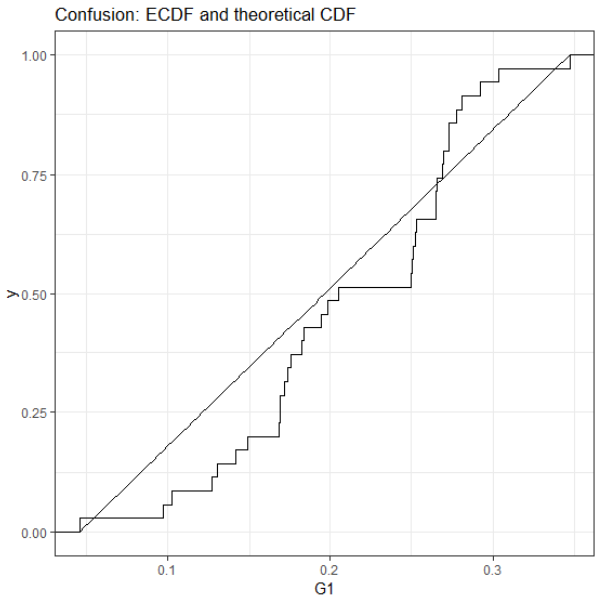


Fig. 3.20 uniform empirical and theoretical CDFs: $conf(p,q)$

In figure 3.20, it is visualized that the staggered line is below and above the adjusted distribution line, so we rejected the uniform distribution.

Experimento 3: Lowest and regular frequency of occurrence of words

For the third experiment, 74 key words were analyzed. These were obtained through a selection by the characteristic of the minimum and regular frequency presented in the historical corpus, with regularization by uniform spread.

In this experiment, 74 of the 111 key words that had the low but regular presence in the corpus were analyzed. As in experiment 2, the values of the quasi-distance measurements were obtained from the relative frequency of the word with respect to its paragraphs (for details of how the relative frequency of the words was calculated see subsection 3.1.3 in page 56 and 57)

The Table 3.2 shown all the values of the similarity functions of the uniform distribution with the respective parameters minimo (min) y máximo (max) , Fitted parameters (min* y max*)and the KS test and its statistics D (statistic with the rank parameter) and D * (statistical of the adjusted parameters) what is the maximum difference between the CDFs of the two corpuses (the D statistic see page 57). All the results are consolidated in this Table 3.2. So, it is important to emphasize that the cosine function is the only one that does not appear as

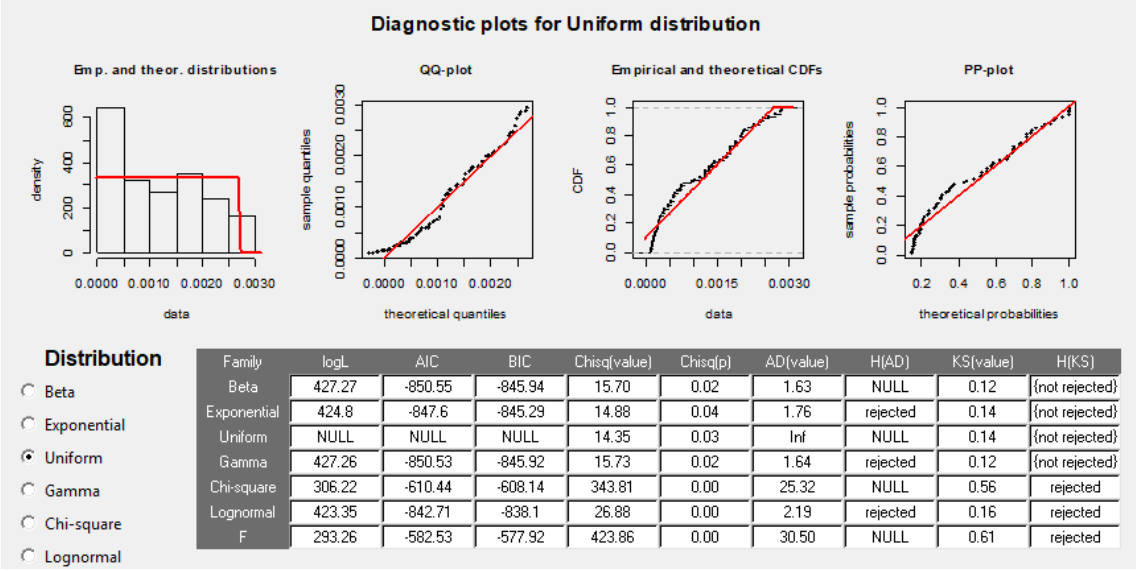
uniformly distributed but rather adjusts better to a beta distribution. Observe the high value of the D * statistic (0.1745)

Table 3.2 **Experiment 3**: Comparison of values functions for probability distributions

Functions	Parameter		Fitted parameter		KS	
	min	max	min*	max*	D	D*
KL(p,q)	0.0001056	0.0029568	-0.0003231	0.0026872	0.2293	0.1424
KL(q,p)	0.000106	0.0029909	-0.0003239	0.0026789	0.2407	0.1431
JS(p,q)	2.649e-05	0.0007461	-8.076e-05	6.690e-04	0.2403	0.1430
JS(q,p)	2.639e-05	0.000738	-8.056e-05	6.711e-04	0.2293	0.1423
$S_\alpha(p,q)$	5.18e-05	0.0014396	0.0001578	0.0013132	0.2315	0.1425
$S_\alpha(q,p)$	5.184e-05	0.0014407	-0.000158	0.0013124	0.2335	0.1426
euc(p,q)	0.0113834	0.073657	0.0082458	0.0636948	0.2151	0.1098
cos(p,q)	0.992689	0.9998193	0.9943993	1.0009651	0.3208	0.1745
$L_1(p,q)$	0.0201541	0.1055499	0.015802	0.1026069	0.1374	0.0934
Conf(p,q)	0.0391235	0.3470588	0.0509106	0.3266327	0.1292	0.1004

Then, we show in detail the figures from 3.21 to 3.40 with their respective interpretations the plots of the uniform distribution and cumulative distribution function (CDF) for each of the similarity indicated in Table 3.2.

a) KL divergence (74 words):$D(p||q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	427.27	-850.55	-845.94	15.70	0.02	1.63	NULL	0.12	{not rejected}
Exponential	424.8	-847.6	-845.29	14.88	0.04	1.76	rejected	0.14	{not rejected}
Uniform	NULL	NULL	NULL	14.35	0.03	Inf	NULL	0.14	{not rejected}
Gamma	427.26	-850.53	-845.92	15.73	0.02	1.64	rejected	0.12	{not rejected}
Chi-square	306.22	-610.44	-608.14	343.81	0.00	25.32	NULL	0.56	rejected
Lognormal	423.35	-842.71	-838.1	26.88	0.00	2.19	rejected	0.16	rejected
F	293.26	-582.53	-577.92	423.86	0.00	30.50	NULL	0.61	rejected

Fig. 3.21 uniform distribution (74 words): $D(p||q)$

Figure 3.21 includes the analysis of the similarity function of the Kullback-Leibler divergence $D(p||q)$, where the first graph represents the density of the uniform distribution. On the second graph the Q-Q plot visually determines that the data is approaching the straight line. On the third graph it is observed that the stepped line follows the adjusted distribution line so the data is adjusted to the distribution and the last graph determines that the data are uniformly distributed, although some discrepancies are observed between the red line and the first one. flat at 0.4 approximately. Also below it presents a table that reflects the numerical values of some of the different goodness of fit tests like: the minor chi-square value of 14.35 that tells us that it fits and where the value of the KS discrepancy is 0.14. These values of the GOF tests assert that the distribution of the $D(p||q)$ fits the uniform probabilistic model.

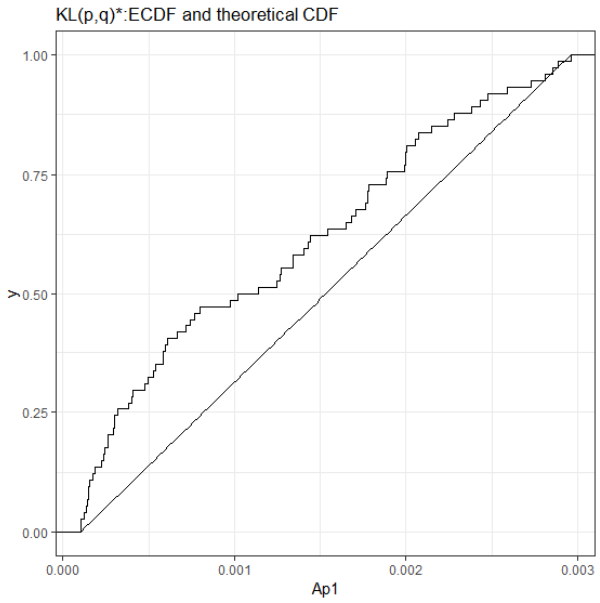
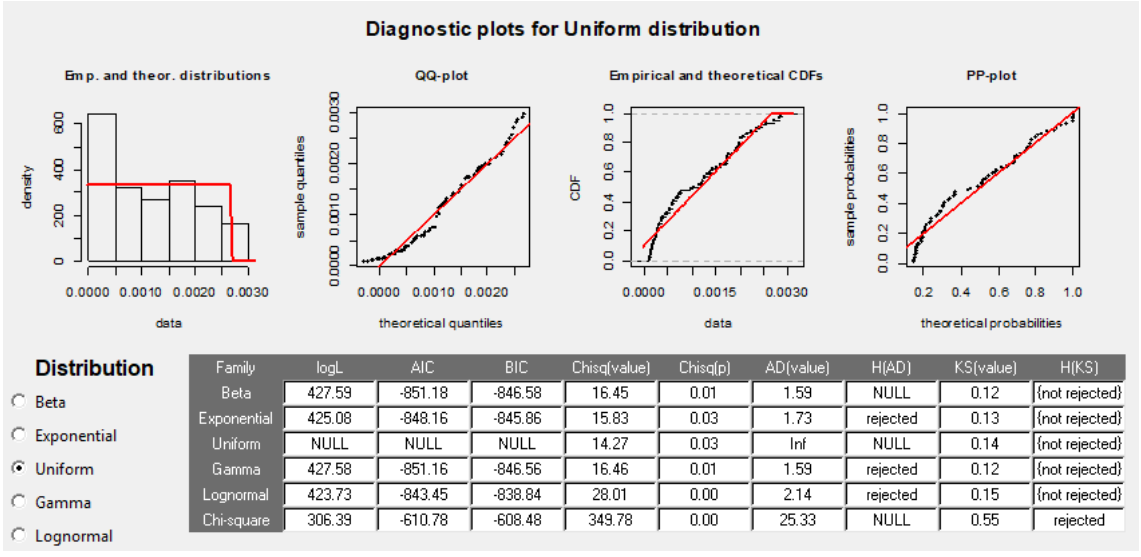


Fig. 3.22 Uniform empirical and theoretical CDFs (74 words): $D(p||q)$

Figure 3.22 shows us that the stepped line is above the adjusted distribution line, so we rejected the uniform distribution.

b) KL divergence (74 words): $D(q||p)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	427.59	-851.18	-846.58	16.45	0.01	1.59	NULL	0.12	{not rejected}
Exponential	425.08	-848.16	-845.86	15.83	0.03	1.73	rejected	0.13	{not rejected}
Uniform	NULL	NULL	NULL	14.27	0.03	Inf	NULL	0.14	{not rejected}
Gamma	427.58	-851.16	-846.56	16.46	0.01	1.59	rejected	0.12	{not rejected}
Lognormal	423.73	-843.45	-838.84	28.01	0.00	2.14	rejected	0.15	{not rejected}
Chi-square	306.39	-610.78	-608.48	349.78	0.00	25.33	NULL	0.55	rejected

Fig. 3.23 uniform distribution (74 words): $D(q||p)$

Figure 3.23 shows the analyzed results of the similarity function of the Kullback leibler divergence $D(q||p)$, the first graph represents the density of the uniform distribution.

On the second graph Q-Q plot reveals that the data approximates a straight line. On the third graph it is observed that the stepped line follows the adjusted distribution line presenting an adequate adjustment and the last graph determines that the data sets are uniformly distributed, although some discrepancies are observed between the red line and the foreground at approximately 0.38 . In turn, below it presents a table that reflects the numerical values of some of the different goodness of fit tests such as: The minor chi-square value of 14.27 that indicates that it fits and the KS value of 0.14. Consequently, these values of the adjustment tests assert that the distribution of the data conforms to the uniform probabilistic model.

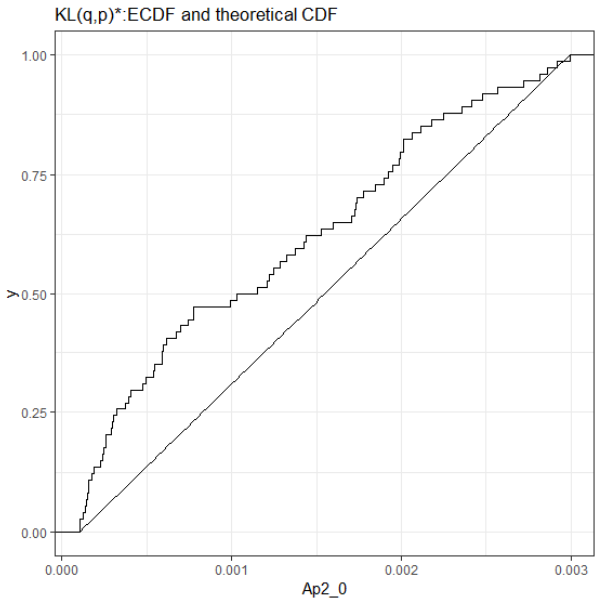
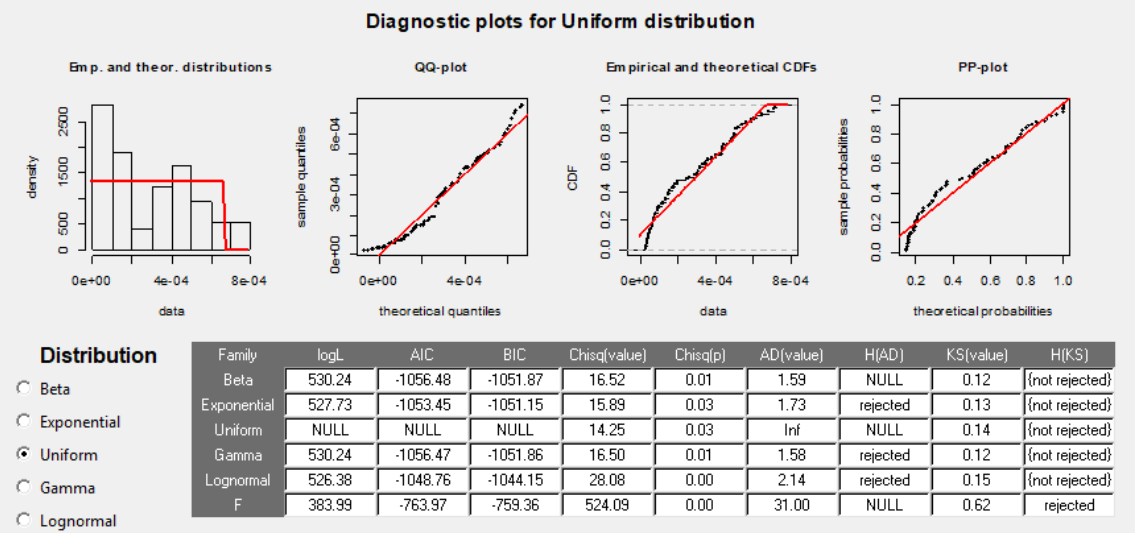


Fig. 3.24 Empirical and theoretical CDFs (74 words): $D(q||p)$

Figure 3.24 shows that the stepped line is above the adjusted distribution line, so it is concluded that the information ranges are not adjusted to the uniform distribution.

c) Jensen-Shannon (74 words):$JS(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	530.24	-1056.48	-1051.87	16.52	0.01	1.59	NULL	0.12	{not rejected}
Exponential	527.73	-1053.45	-1051.15	15.89	0.03	1.73	rejected	0.13	{not rejected}
Uniform	NULL	NULL	NULL	14.25	0.03	Inf	NULL	0.14	{not rejected}
Gamma	530.24	-1056.47	-1051.86	16.50	0.01	1.58	rejected	0.12	{not rejected}
Lognormal	526.38	-1048.76	-1044.15	28.08	0.00	2.14	rejected	0.15	{not rejected}
F	383.99	-763.97	-759.36	524.09	0.00	31.00	NULL	0.62	rejected

Fig. 3.25 uniform distribution (74 words): $JS(p,q)$

Figure 3.25 shows the analysis of the Jensen-Shannon similarity function $JS(p,q)$ exhibits a first graph that presents the density of the uniform distribution. On the second graph Q-Qplot represents how the data approaches the straight line. On the third graph it is observed that the staggered line follows the adjusted distribution line, so the data fits properly to the distribution and the last graph determines that the data set is uniformly distributed although some discrepancy is observed between the red line with the foreground at 0.4 approximately. Also, the presence of the table below shows the numerical values of some of the different goodness of fit tests as: the lower chi square value of 14.25 indicates that it is credible and the KS value of 0.14. All of these GOF test values determine that the distributions of the data conform with the uniform probability distribution.

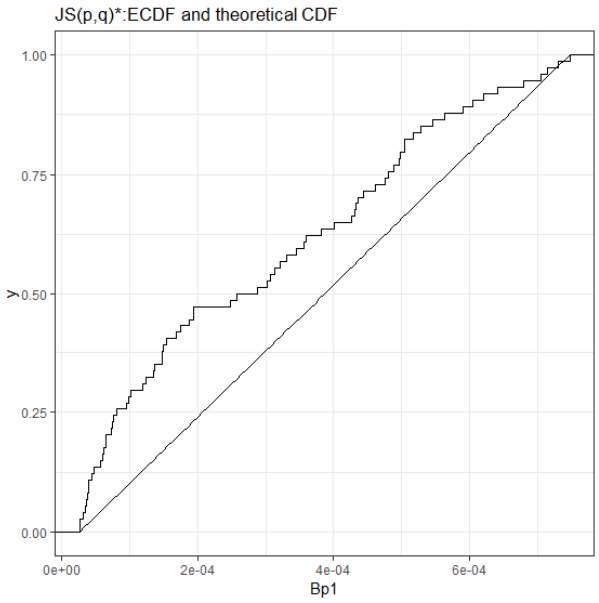
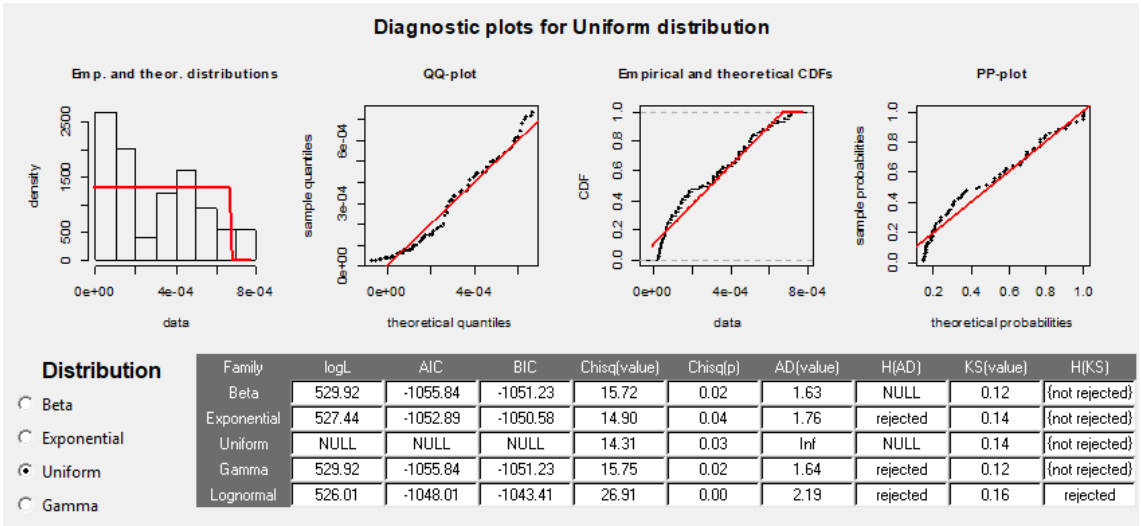


Fig. 3.26 Empirical and theoretical (74 words): $JS(p,q)$

Figure 3.26 shows that the stepped line is on the adjusted distribution line, so it is concluded that the information ranges are deficient adjusted to the uniform distribution.

d) Jensen-Shannon (74 words):$JS(q,p)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	529.92	-1055.84	-1051.23	15.72	0.02	1.63	NULL	0.12	{not rejected}
Exponential	527.44	-1052.89	-1050.58	14.90	0.04	1.76	rejected	0.14	{not rejected}
Uniform	NULL	NULL	NULL	14.31	0.03	Inf	NULL	0.14	{not rejected}
Gamma	529.92	-1055.84	-1051.23	15.75	0.02	1.64	rejected	0.12	{not rejected}
Lognormal	526.01	-1048.01	-1043.41	26.91	0.00	2.19	rejected	0.16	rejected

Fig. 3.27 uniform distribution (74 words): $JS(q,p)$

Figure 3.27 presents the analysis of the similarity function of Jensen-Shannon $JS(q,p)$ where the first graph shows the density of the uniform distribution. On the second

graph Q-Qplot visually determines that the data is close to the line. On the third graph it is observed that the stepped line follows the line of the adjusted distribution, so that the data adjust appropriately to the distribution and the last graph determines that the data present a uniform behavior although some discrepancies are observed between the red line with the foreground at 0.4 approximately. Also, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: the smaller chi-square value of 14.31 indicates that it fits the distribution and the KS value of 0.14. These values of GOF-tests determine that the distribution of the data fits the uniform probability distribution.

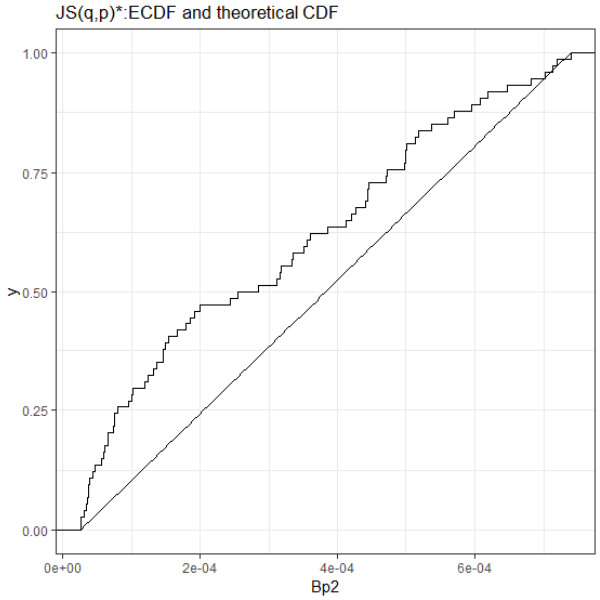
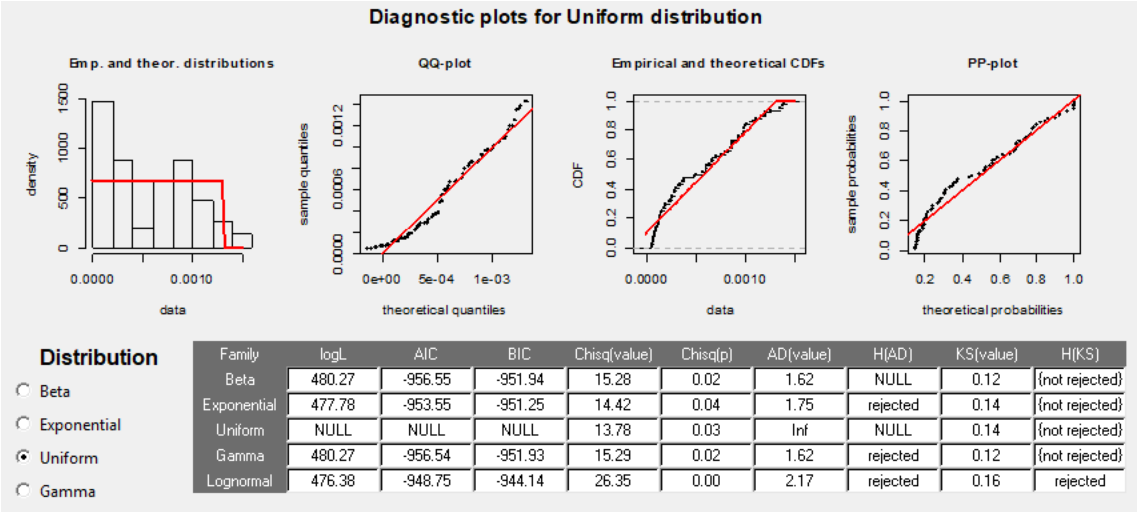


Fig. 3.28 Empirical and theoretical CDFs (74 words): $JS(q,p)$

In Figure 3.28, it is observed that the staggered line is on the adjusted distribution line so we rejected the uniform distribution.

e) Skew divergence (74 words):$S_\alpha(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	480.27	-956.55	-951.94	15.28	0.02	1.62	NULL	0.12	{not rejected}
Exponential	477.78	-953.55	-951.25	14.42	0.04	1.75	rejected	0.14	{not rejected}
Uniform	NULL	NULL	NULL	13.78	0.03	Inf	NULL	0.14	{not rejected}
Gamma	480.27	-956.54	-951.93	15.29	0.02	1.62	rejected	0.12	{not rejected}
Lognormal	476.38	-948.75	-944.14	26.35	0.00	2.17	rejected	0.16	rejected

Fig. 3.29 uniform distribution (74 words): $S_\alpha(p,q)$

Figure 3.29 shows the analysis of the similarity function of Skew divergence $S_\alpha(p,q)$ here the first graph shows the density of the uniform distribution. On the second graph Q-Q plot it is visualized that the data approaches the straight line. On the third graph it is observed that the stepped line follows the adjusted distribution line so that the data fits properly to the distribution and the last graph determines that the data set is uniformly distributed although some discrepancies are observed between the red line and the foreground at 0.4 approximately. In turn, it presents a table in the lower part where the numerical values of some of the different goodness of fit tests are observed as: the smaller chi-square value of 13.78 indicates that it fits the distribution and the KS value of 0.14. These values of GOF-tests determine that the distribution of the data fits the uniform probability distribution.

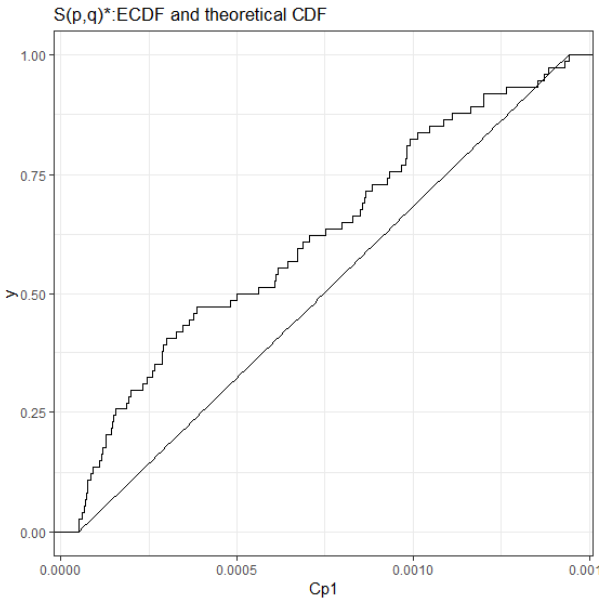
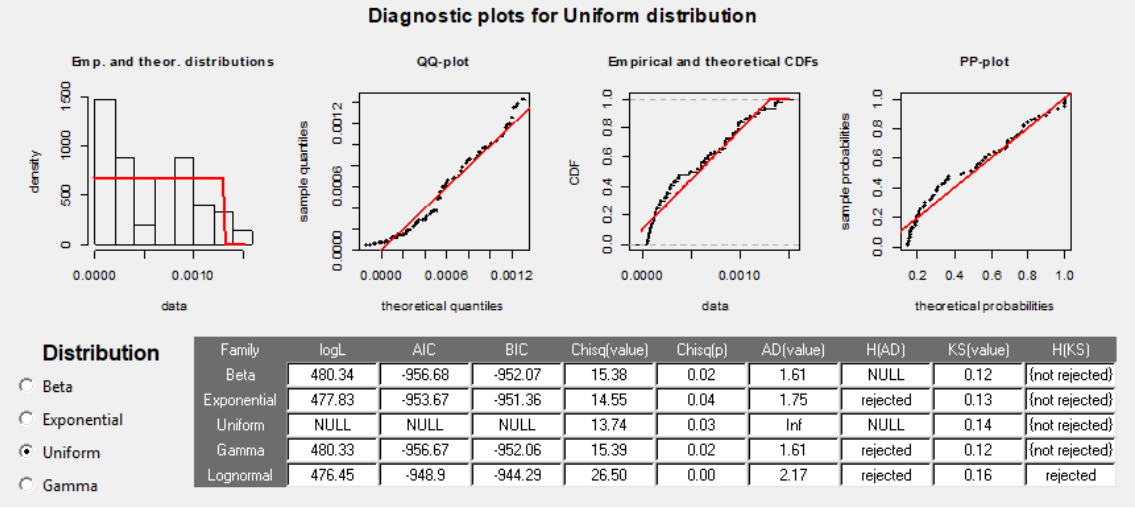


Fig. 3.30 Empirical and theoretical (74 words): $S_\alpha(p,q)$

Figure 3.30, allows us to observe a staggered line that is on the adjusted distribution line so we rejected the uniform distribution.

f) Skew divergence (74 words):$S_\alpha(q,p)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	480.34	-956.68	-952.07	15.38	0.02	1.61	NULL	0.12	{not rejected}
Exponential	477.83	-953.67	-951.36	14.55	0.04	1.75	rejected	0.13	{not rejected}
Uniform	NULL	NULL	NULL	13.74	0.03	Inf	NULL	0.14	{not rejected}
Gamma	480.33	-956.67	-952.06	15.39	0.02	1.61	rejected	0.12	{not rejected}
Lognormal	476.45	-948.9	-944.29	26.50	0.00	2.17	rejected	0.16	rejected

Fig. 3.31 uniform distribution (74 words): $S_\alpha(q,p)$

Figure 3.31 describes the analysis of the similarity function of Skew divergence $S_\alpha(q,p)$ the first graph shows the density of the uniform distribution. On the second

graph Q-Q plot allows us to observe that the data is almost next to the straight line. On the third graph we perceive that the staggered line follows the adjusted distribution line so the data fit appropriately to the distribution and the last graph shows that the data show a uniform behavior although some discrepancies are observed between the red line and the first flat at 0.4 approximately. Likewise, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: the smaller chi-square value of 13.74 indicates that it fits the distribution and the KS value of 0.14. These values of GOF-tests allow us to assert that the distributions of the data fit the uniform probability distribution.

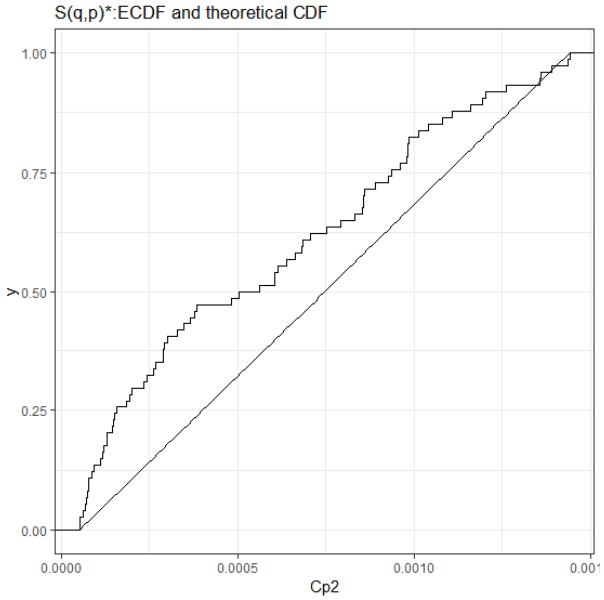
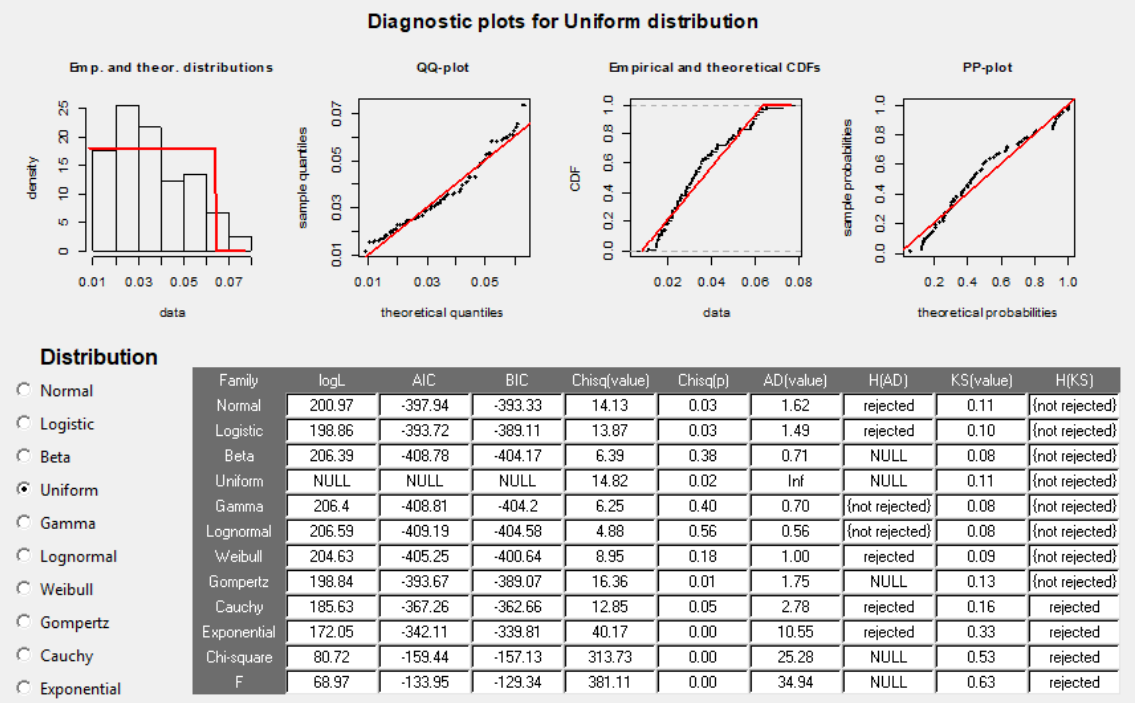


Fig. 3.32 Empirical and theoretical (74 words): $S_\alpha(q,p)$

Figure 3.32 shows how the staggered line is on the adjusted distribution line, so we rejected the uniform distribution.

g) Euclidean (74 words):$euc(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	200.97	-397.94	-393.33	14.13	0.03	1.62	rejected	0.11	{not rejected}
Logistic	198.86	-393.72	-389.11	13.87	0.03	1.49	rejected	0.10	{not rejected}
Beta	206.39	-408.78	-404.17	6.39	0.38	0.71	NULL	0.08	{not rejected}
Uniform	NULL	NULL	NULL	14.82	0.02	Inf	NULL	0.11	{not rejected}
Gamma	206.4	-408.81	-404.2	6.25	0.40	0.70	{not rejected}	0.08	{not rejected}
Lognormal	206.59	-409.19	-404.58	4.88	0.56	0.56	{not rejected}	0.08	{not rejected}
Weibull	204.63	-405.25	-400.64	8.95	0.18	1.00	rejected	0.09	{not rejected}
Gompertz	198.84	-393.67	-389.07	16.36	0.01	1.75	NULL	0.13	{not rejected}
Cauchy	185.63	-367.26	-362.66	12.85	0.05	2.78	rejected	0.16	rejected
Exponential	172.05	-342.11	-339.81	40.17	0.00	10.55	rejected	0.33	rejected
Chi-square	80.72	-159.44	-157.13	313.73	0.00	25.28	NULL	0.53	rejected
F	68.97	-133.95	-129.34	381.11	0.00	34.94	NULL	0.63	rejected

Fig. 3.33 uniform distribution (74 words): $euc(p,q)$

Figure 3.33 shows the analysis of the similarity function of the Euclidean $euc(p,q)$ where the first graph shows the density of the uniform distribution. On the second graph Q-Q plot visually determines that the data is closer to the straight line although there are outliers. On the third graph we perceive that the stepped line follows the adjusted distribution line so that the data fit appropriately to the distribution and the last graph shows that the data present a uniform behavior, although some discrepancies are observed between the red line and the close-up on 0.5 and 0.9 approximately. In turn, it presents a table in the lower part where the numerical values of some different goodness of fit tests are observed as: The lower chi square value of 14.82 indicates that it fits the distribution and the KS value of 0.11. These values of GOF-tests confirm that the data fit the uniform probability distribution.

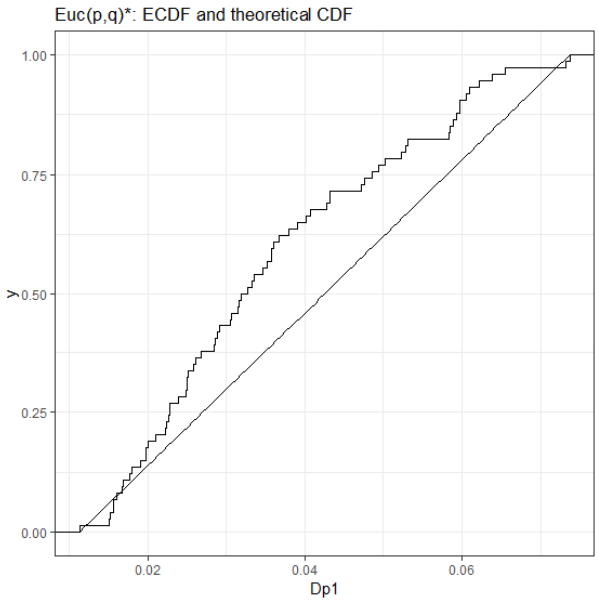
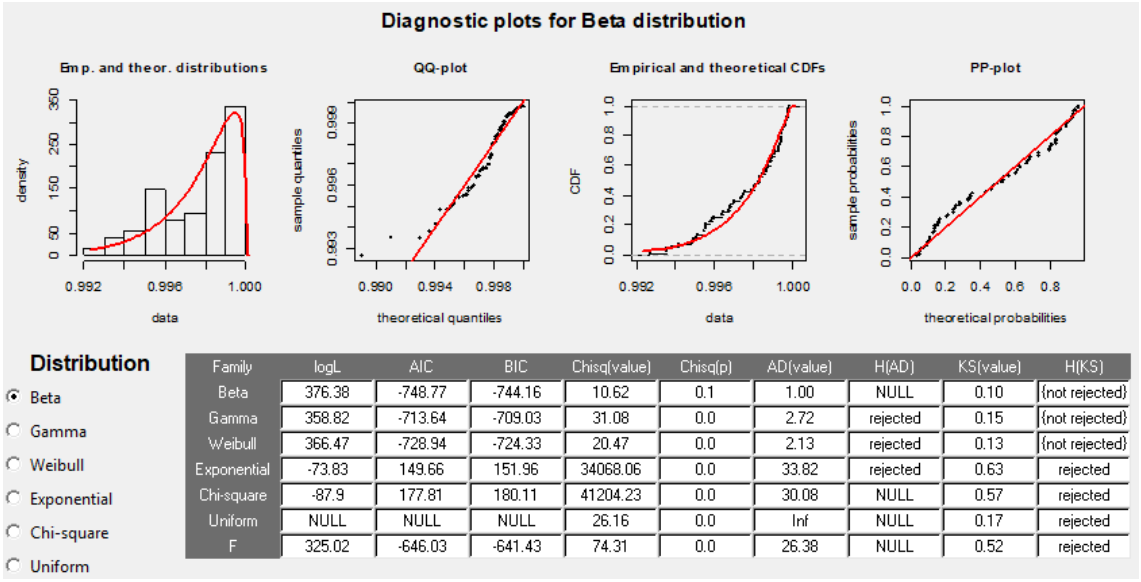


Fig. 3.34 Empirical and theoretical CDFs (74 words): $euc(p,q)$

Figure 3.34, allows us to observe that the staggered line is on the adjusted distribution line so we rejected the uniform distribution.

h) Cosine (74 words):$cos(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Beta	376.38	-748.77	-744.16	10.62	0.1	1.00	NULL	0.10	{not rejected}
Gamma	358.82	-713.64	-709.03	31.08	0.0	2.72	rejected	0.15	{not rejected}
Weibull	366.47	-728.94	-724.33	20.47	0.0	2.13	rejected	0.13	{not rejected}
Exponential	-73.83	149.66	151.96	34068.06	0.0	33.82	rejected	0.63	rejected
Chi-square	-87.9	177.81	180.11	41204.23	0.0	30.08	NULL	0.57	rejected
Uniform	NULL	NULL	NULL	26.16	0.0	Inf	NULL	0.17	rejected
F	325.02	-646.03	-641.43	74.31	0.0	26.38	NULL	0.52	rejected

Fig. 3.35 beta distribution (74 words): $cos(p,q)$

Figure 3.35 shows the analysis of the cosine similarity function. Where the 4 graphs (density distribution, QQ plot, CDFs and PP-plot) of the $cos(p,q)$ and the table that presents in the lower part with the numerical values of some goodness of fit tests like: the value of chi square and the discriminant value of KS affirm that the data do not behave in a uniform manner but show a beta behavior.

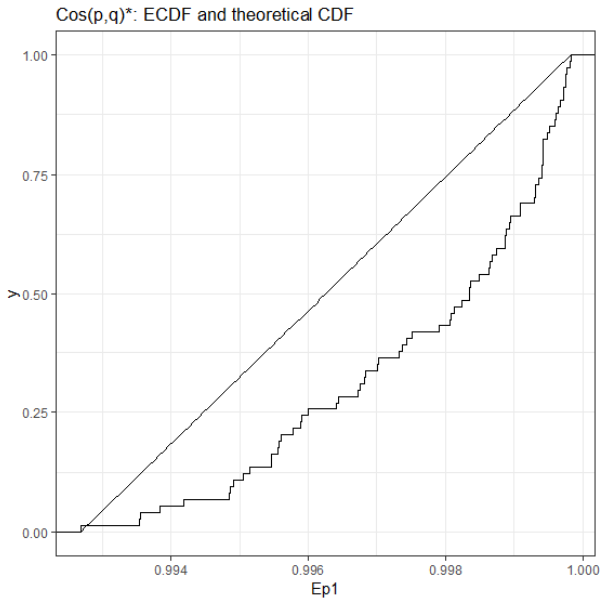
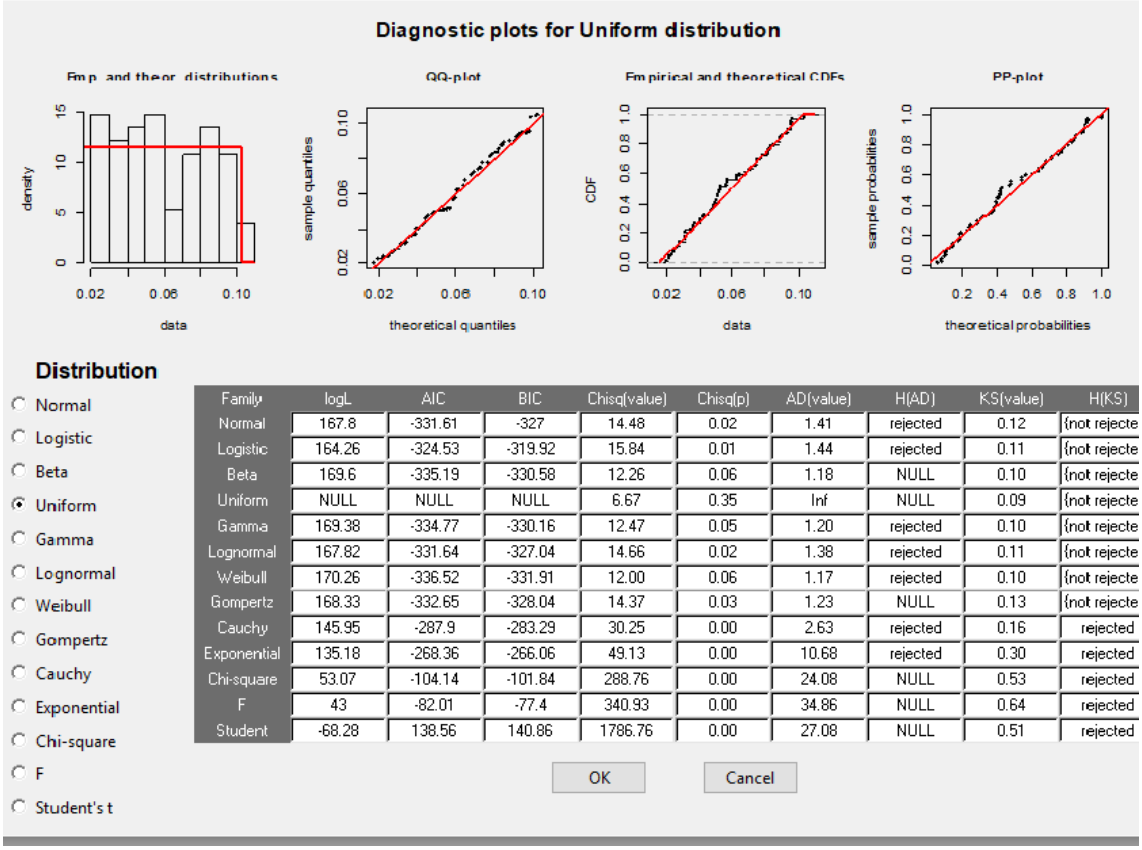


Fig. 3.36 Empirical and theoretical CDFs (74 words): $cos(p,q)$

In Figure 3.36, it is perceived that the stepped line is well below the adjusted distribution line, so it is concluded that the information ranges do not fit to the uniform distribution.

i) L_1(74 words): $L_1(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	167.8	-331.61	-327	14.48	0.02	1.41	rejected	0.12	{not rejecte
Logistic	164.26	-324.53	-319.92	15.84	0.01	1.44	rejected	0.11	{not rejecte
Beta	169.6	-335.19	-330.58	12.26	0.06	1.18	NULL	0.10	{not rejecte
Uniform	NULL	NULL	NULL	6.67	0.35	Inf	NULL	0.09	{not rejecte
Gamma	169.38	-334.77	-330.16	12.47	0.05	1.20	rejected	0.10	{not rejecte
Lognormal	167.82	-331.64	-327.04	14.66	0.02	1.38	rejected	0.11	{not rejecte
Weibull	170.26	-336.52	-331.91	12.00	0.06	1.17	rejected	0.10	{not rejecte
Gompertz	168.33	-332.65	-328.04	14.37	0.03	1.23	NULL	0.13	{not rejecte
Cauchy	145.95	-287.9	-283.29	30.25	0.00	2.63	rejected	0.16	rejected
Exponential	135.18	-268.36	-266.06	49.13	0.00	10.68	rejected	0.30	rejected
Chi-square	53.07	-104.14	-101.84	288.76	0.00	24.08	NULL	0.53	rejected
F	43	-82.01	-77.4	340.93	0.00	34.86	NULL	0.64	rejected
Student	-68.28	138.56	140.86	1786.76	0.00	27.08	NULL	0.51	rejected

Fig. 3.37 uniform distribution (74 words): $L_1(p,q)$

Figure 3.37 presents the analysis of the similarity function L_1, where the first graph of $L_1(p,q)$ displays the density of the uniform distribution. On the second graph Q-Qplot visually determines that the data is almost next to the line. On the third graph it is observed that the staggered line closely follows the adjusted distribution line so that the data fit appropriately to the distribution and the last graph determines that the data sets show a uniform behavior, although some discrepancies are observed between the red line and the first plane at 0.5 approximately. Also, it presents a table in the lower part where the numerical values of some goodness of fit tests are observed, such as: the smaller chi-square value of 6.67 that indicates that it fits the distribution, as well as the KS value of 0.09, confirm that the data They have a uniform distribution.

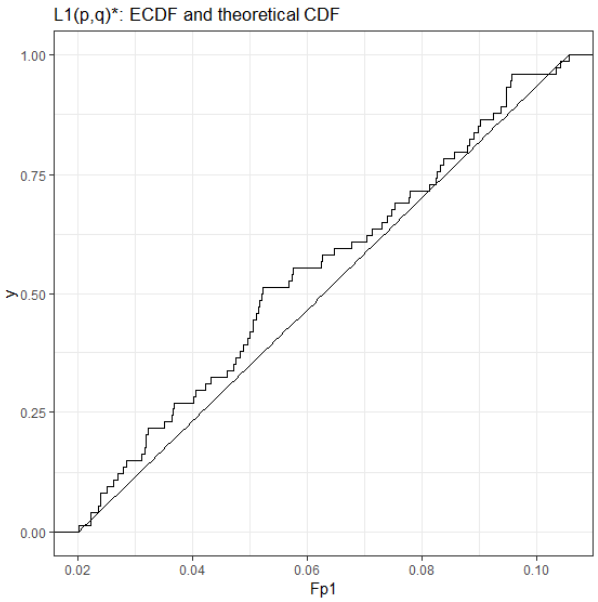
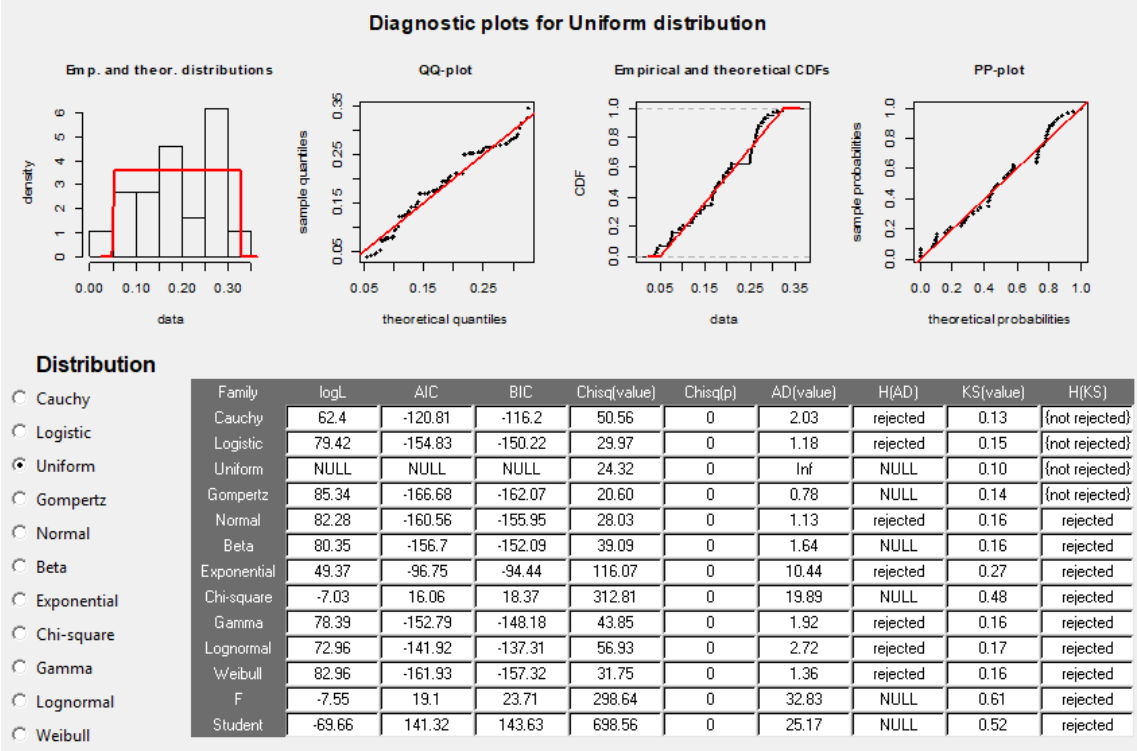


Fig. 3.38 Uniform empirical and theoretical CDFs (74 words): $L_1(p,q)$

In Figure 3.38, it is shown that the stepped line follows the adjusted distribution line, so it is affirmed that the information ranges fit almost uniformly to the uniform distribution.

j) Confusion (74 words):$conf(p,q)$



Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Cauchy	62.4	-120.81	-116.2	50.56	0	2.03	rejected	0.13	{not rejected}
Logistic	79.42	-154.83	-150.22	29.97	0	1.18	rejected	0.15	{not rejected}
Uniform	NULL	NULL	NULL	24.32	0	Inf	NULL	0.10	{not rejected}
Gompertz	85.34	-166.68	-162.07	20.60	0	0.78	NULL	0.14	{not rejected}
Normal	82.28	-160.56	-155.95	28.03	0	1.13	rejected	0.16	rejected
Beta	80.35	-156.7	-152.09	39.09	0	1.64	NULL	0.16	rejected
Exponential	49.37	-96.75	-94.44	116.07	0	10.44	rejected	0.27	rejected
Chi-square	-7.03	16.06	18.37	312.81	0	19.89	NULL	0.48	rejected
Gamma	78.39	-152.79	-148.18	43.85	0	1.92	rejected	0.16	rejected
Lognormal	72.96	-141.92	-137.31	56.93	0	2.72	rejected	0.17	rejected
Weibull	82.96	-161.93	-157.32	31.75	0	1.36	rejected	0.16	rejected
F	-7.55	19.1	23.71	298.64	0	32.83	NULL	0.61	rejected
Student	-69.66	141.32	143.63	698.56	0	25.17	NULL	0.52	rejected

Fig. 3.39 uniform distribution (74 words): $conf(p,q)$

Figure 3.39 shows the analysis of the Confusion similarity function. Where the first graph of $conf(p,q)$ displays the density of the uniform distribution. On the second graph Q-Q plot visually determines that the data is close to the straight line. On the third graph we perceive that the stepped line follows the adjusted distribution line so the data is adjusted to the distribution and the last graph shows that the data have a Uniform behavior although some discrepancies are observed in the red line and the foreground in 0.7 approximately. In turn, it presents a table in the lower part where the numerical values of some goodness of fit tests are observed, such as: the smaller chi-square value of 24.32 indicates that it fits the distribution and the KS value of 0.10. These values of GOF-tests determine that the data have a uniform distribution.

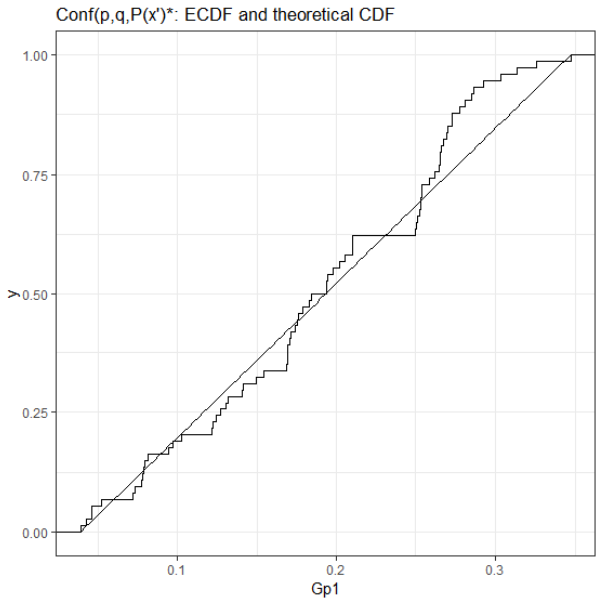


Fig. 3.40 Empirical and theoretical (74 words): $conf(p,q)$

In figure 3.40, it is visualized that the stepped line follows almost closely to the adjusted distribution line, so it is affirmed that the ranges of the information fit almost uniformly to the uniform distribution.

In the next section we digress to the study of the uniform design of Fang, Kai-Tai et al.(2000) Fang et al. (2000b), in order to illustrate importance of uniform approximation of quasi-distances.

3.2 Uniform designs: measures of Uniformity

We introduce the uniform designs in order to better emphasize their importance in information theory without prior knowledge about informative events, namely letters and key words of Nahuatl language. This is important because we need to learn subsets of letters and key words of Nahuatl for wich we obtain equally distributed quasi-distance betwen both corpuses.

(Fang et al., 2000a) The uniform design (UD) seeks design points that are uniformly scattered on the domain. Finding that the UD's have many desirable properties for a wide variety of applications. Use the global optimization algorithm, threshold sccept-

ing, to generate UD's with low discrepancy. Here investigated the relationship between uniformity and orthogonality . It turns out that most UD's obtained here are indeed ortogonal. The principal idea of UD is that one should choose a set of experimental points with smallest discrepancy among all possible designs for a given number of factors and experimental runs.

Let s factors of interest over a standard domain C^s. The goal here is to choose a set of n points $P_n = \{x_1, ..., x_n\} \subset C^s$ such that these points are uniformly scattered on C^s. Let $M(P_n)$ be a measure of the nonuniformity of P_n. Fang et al.(2000) seek a set P_n^* that minimizes M or, equivalently, maximizes the uniformity over all possible n points on C^s.

A natural choice of M is the discrepancy $D(p)$. Let $F_n(x)$ be the empirical distribution function of P_n :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{x_i \leq x\} \tag{3.5}$$

where $I\{.\}$ is the indicator function and all inequalities are understood to be with respect to the componentwise order of R^s. Then the L_p discrepancy be defined as

$$D_p(P_n) = [\int_{C^s} |F_n(x) - F(x)|^p dx]^{1/p} \tag{3.6}$$

where $F(x)$ is the uniform distribution function on C^s. The popular L_∞ discrepancy obtained by taking $p = \infty$ is called the star discrepancy, or discrepancy for simplicity. This is probably the most commonly used measurement for discrepancy and can be reexpressed as follows:

$$D(P_n) = sup_{x \in C^s} |F_n(x) - F(x)| \tag{3.7}$$

The discrepancy has been universally accepted in quasi Monte Carlo methods and number-theoretic methods. In fact, the discrepancy is the Smirnov-Kolmogorov statistic for goodness-of-fit tests. One disvantage of the discrepancy is that it is expensive to compute. Attempts have been made to evaluate the discrepancy algorithmically.

So, $D(P_n)$ is the largest absolute difference observed between the empirical distribution function of P_n and the uniform distribution function on C^s, obtained from the distribution of probability that is specified as a null hypothesis. If $F_n(x)$ are similar to $F(x)$, the

value of $D(P_n)$ will be small.That is, minimum discrepancy. Also the researchers find that a UD generated by a U-type design has lower discrepancy. Moreover the uniform design es D-optimal. Finally, they come to the conjecture that uniform designs for a suitable measure of uniformity will be orthogonal.

Chapter 4

Conclusions

4.1 Formal statistics for the pseudo-distances between two linguistic corpuses

In this thesis we have suggested a more efficient and more reliable measures of pseudo-distances or similarity to measure the frequencies of letters and words between the corpuses. The pseudo-distances between two linguistic corpuses can explain the changes in the text suffered during the evolution of time. Such changes are intrinsically related to the spelling and semantics of the language. According to our empirical results obtained from the application of similarity methods of probability distribution (KL-divergence, Jensen-Shannon divergence, skew divergence Euclidean, Cosine, L_1 and Confusion probability) the asymmetry is not negligible in such comparisons between words of two different corpuses of Nican Mopohua (Nahuat-Mexico Report). Therefore, we can conjecture that the phenomenon of linguistic variability is asymmetric (see also Discussion paper Pigoli et al. (2018) (Stehlík and Pari, 2018)), which underlines the importance of topological (sometimes non metrizable) approach to linguistic information aggregation (see Stehlík (2016), and tables 2.2 and 2.3 in chapter 2). We also illustrated construction of topological neighbourhoods of some key words. KL-divergence yields best results from all considered pseudo-distances.Because, KL-divergence has good information properties which are very useful for computation of quasi-distances between both corpuses. Namely, this measure is related to robust properties. The language Náhuatl has an agglutinating nature and a complex structure formed by lexemes, morphemes and affixes (neighbouring structures). As any other language it varies in time. The study analyzes the language of classic Nahuatl,

which still has a great number of speakers, namely 1,725,000 inhabitants (INAI). We acknowledge the kind direct support of Nahuatl Aboriginal teachers of Mexico City (House of Culture of Azcapotzalco of the DF of Mexico, Department of Studies in Indigenous Languages of the CUCSH University of Guadalajará and National Autonomous University of Mexico) that helped us to determine the alphabet (20 between letters and digraphs) and the translation of the key words (111 words). We analyzed the relative frequencies obtained from the alphabet and the key words obtained from the comparison of the two corpuses of Lasso (1649) and Rojas (1978).
Further investigation in order to determine if the language changes in spelling and semantics are more related to cultural or to linguistic evolution, e.g. by application the semantic change measures of the Global and Local neighborhood (Hamilton W., 2016) will be a valuable future research direction.

4.2 Uniform distribution of pseudo-distance

We found that uniform distribution provides good fit for pseudo-distances in the second experiment consisting of 35 words that presented lower frequency of occurrence (see Figures 3.1 to 3.20 in Chapter 3). It should be recalled that the uniform distribution is the maximum entropy distribution on any interval [a, b]. We can consider it to be a good choice as uninformative prior for cases where we are lacking additional sources of information. This can be interpreted by meaning that all possible values are equally likely a priori, or you have no prior information.
(Fang, 2000) (Fang et al., 2000a) investigate the uniform design (UD) i.e. design points that are uniformly scattered on the domain. The main idea of UD is that one should choose a set of experimental points with smallest discrepancy among all possible designs for a given number of factors and experimental runs.
In fact, this discrepancy is the Smirnov-Kolmogorov statistic for goodness-of-fit tests. One disvantage of the discrepancy is that it is expensive to compute. It is convenient to obtain a minimum discrepancy (see Table 3.1). Also the researchers find that a UD generated by a U-type design has lower discrepancy. Moreover the uniform design is D-optimal for estimation of trend parameter of Ornstein-Uhlenbeck process (see Kisel'ák and Stehlík (2008)). In addition, uniform distribution provides a good fit to pseudo-distances of given words form two corpuses of Nican Mopohua in third experiment with 74 words of minor and regular frequency (see Figure 3.21 to 3.40). Table 3.2 displays the results of the different tests of goodness of fit for the last experiment of 74 words.

4.3 Generalized gamma distribution of pseudo-distance

Naturally uniform distribution can fail to provide a good fit for pseudo-distances between both corpuses of Nican Mopohua for general subset of words, e.g. by considering the first experiment with an analysis of all the key words (111 words) of the Nican Mopohua. Since we consider non-negative pseudo-distances, generalized gamma distribution provides a reasonable fit. This is justified both by theoretical results cited in the thesis, also by our empirical data experiments. So we conjecture that the empirical process of pseudo-distances can be well approximated by generalized gamma distribution as a suitable model of semantic changes between two corpuses of Nican Mopohua.

Likewise, the p-value and the values of the other goodness-of-fit measures (logL, AIC, BIC, Chi squared, AD and KS) of the psuedo-distances confirm the choice of such model (see figure 2.8 to 2.17). Thus we also conjecture approximation by gamma distribution. We should also admit that distance measures Cosine and Confusion are violating such kind of approximation.

References

R. Amato, L. Lacasa, A. Díaz-Guilera, and A. Baronchelli. The dynamics of norm change in the cultural evolution of language. *Proceedings of the National Academy of Sciences*, 115(33):8260–8265, 2018.

N. Belgorodski, M. Greiner, K. Tolksdorf, K. Schueller, M. Flor, L. Göhring, and M. M. Greiner. Package 'rriskdistributions'. *Online: https://cran. r-project. org/web/packages/rriskDistributions/rriskDistributions. pdf (08.11. 2018)*, 2017.

B. Bigi. Using kullback-leibler distance for text categorization. In *European Conference on Information Retrieval*, pages 305–319. Springer, 2003.

L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

I. CSISZAR. Eine informations theoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten. *Mayar Tud. Acad. Mat. Kutato Int. Kozl.*, 8:85–108, 1963.

I. Dagan, L. Lee, and F. C. Pereira. Similarity-based models of word cooccurrence probabilities. *Machine learning*, 34(1-3):43–69, 1999.

G. M. Ortiz. *Nican Mopouha*. Departamento de Ciencias Religiosas, Universidad Iberoamericana, 1990.

U. Essen and V. Steinbiss. Cooccurrence smoothing for stochastic language modeling. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 161–164. IEEE, 1992.

K.-T. Fang, D. K. Lin, P. Winker, and Y. Zhang. Uniform design: theory and application. *Technometrics*, 42(3):237–248, 2000a.

K.-T. Fang, D. K. Lin, P. Winker, and Y. Zhang. Uniform design: theory and application. *Technometrics*, 42(3):237–248, 2000b.

R. A. Fisher. On the interpretation of χ 2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.

R. A. Fisher. Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 22, pages 700–725. Cambridge University Press, 1925.

A. M. Garibay. La have del nahuatl. *Colección de Trozos Clasicos, etc. México*, pages 1953–54, 1940.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*, volume 127. Cambridge University Press, 2009.

S. Greenland. Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values. *The American Statistician*, 73(sup1): 106–114, 2019.

I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley. Analysis of symbolic sequences using the jensen-shannon divergence. *Physical Review E*, 65(4):041905, 2002.

R. B. Gutierrez and P. G. i Cintas. El histograma como un instrumento para la comprensión de las funciones de densidad de probabilidad. *Probabilidad Condicionada: Revista de didáctica de la Estadística*, (2):229–235, 2013.

M. Hilpert and S. T. Gries. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4):385–401, 2008.

H. Hubey. Mathematical methods in historical linguistics; their use, misuse and abuse. *submitted to Journal of the IQLA*, 1999.

C. Illert and A. Allison. Phono-genesis and the origin of accusative syntax in proto-australian language. *Journal of Applied Statistics*, 31(1):73–104, 2004.

T. Kanamori and M. Sugiyama. Statistical analysis of distance estimators with density differences and density ratios. *Entropy*, 16(2):921–942, 2014a.

T. Kanamori and M. Sugiyama. Statistical analysis of distance estimators with density differences and density ratios. *Entropy*, 16(2):921–942, 2014b.

D. F. Kerridge. Inaccuracy and inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 184–194, 1961.

J. Kisel'ák and M. Stehlík. Equidistant and d-optimal designs for parameters of ornstein–uhlenbeck process. *Statistics & Probability Letters*, 78(12):1388–1396, 2008.

F. C. Klebaner et al. Stochastic difference equations and generalized gamma distributions. *The Annals of Probability*, 17(1):178–188, 1989.

S. Kullback. *Information theory and statistics*. Courier Corporation, 1997.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. ISSN 00034851. URL http://www.jstor.org/stable/2236703.

L. Lasso De La Vega. 1649. *Huei tlamahuicoltica omonexiti in ilhuicac tlatocacihuapilli Santa Maria Totlaconantzin Guadalupe in nican huei altepenahuac Mexico itocayocan tepeyacac*.

L. Lee. Measures of distributional similarity. *arXiv preprint cs/0001012*, 2000.

L. Lee. On the effectiveness of the skew divergence for statistical language analysis. In *AISTATS*. Citeseer, 2001.

F. Liese and I. Vajda. Convex statistical distances. 1987.

J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

D. V. Lindley et al. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

B. G. Lindsay et al. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *The annals of statistics*, 22(2):1081–1114, 1994.

F. Makri, Z. Psillakis, et al. On limited length binary strings with an application in statistical control. *The Open Statistics and Probability Journal*, 8(1):1–6, 2017.

M. Menéndez, D. Morales, L. Pardo, and I. Vajda. Two approaches to grouping of data and related disparity statistics. *Communications in Statistics-Theory and Methods*, 27(3):609–633, 1998.

J. Montero, D. Gómez, V. López, J. T. Rodríguez, and B. Vitoriano. Sobre funciones y reglas de agregación.

D. Morales, L. Pardo, and I. Vajda. Divergence between various estimates of quantized information sources. *Kybernetika*, 32(4):395–407, 1996.

A. H. Murphy. A new vector partition of the probability score. *Journal of applied Meteorology*, 12(4):595–600, 1973.

L. Pardo. *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2005.

M. Pardo and I. Vajda. About distances of discrete distributions satisfying the data processing theorem of information theory. *IEEE transactions on information theory*, 43(4):1288–1293, 1997.

K. Pearson. On the theory of contingency and its relation to association and normal correlation, biometric series no. 1. *Drapers' Co. Memoirs, London*, 1904.

D. Pigoli, P. Z. Hadjipantelis, J. S. Coleman, and J. A. Aston. The statistical analysis of acoustic phonetic data: exploring differences between spoken romance languages. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1103–1145, 2018.

S. Rojas. Mario, traducción del náhuatl al castellano del texto nican mopohua. v. valeriano, antonio. *Nican Mopohua*, 1978.

C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

M. Stehlík. Distributions of exact tests in the exponential family. *Metrika*, 57(2): 145–164, 2003.

M. Stehlík. Homogeneity and scale testing of generalized gamma distribution. *Reliability Engineering & System Safety*, 93(12):1809–1813, 2008.

M. Stehlík. On convergence of topological aggregation functions. *Fuzzy Sets and Systems*, 287:48–56, 2016.

M. Stehlík and M. Pari. Discussion on" the statistical analysis of acoustic phonetic data: exploring differences between spoken romance languages". 2018.

S. M. Stigler. *The seven pillars of statistical wisdom*. Harvard University Press, 2016.

J. A. Villaseñor and E. González-Estrada. A variance ratio test of fit for gamma distributions. *Statistics & Probability Letters*, 96:281–286, 2015.

R. L. Wasserstein, A. L. Schirm, and N. A. Lazar. Moving to a world beyond "p< 0.05", 2019.

S. V. Weijs, R. Van Nooijen, and N. Van De Giesen. Kullback–leibler divergence as a forecast skill score with classic reliability–resolution–uncertainty decomposition. *Monthly Weather Review*, 138(9):3387–3399, 2010.

C. P. William. Minimum distance estimation: a bibliography. *Communications in Statistics-Theory and Methods*, 10(12):1205–1224, 1981.

F. Yates. Contingency tables involving small numbers and the χ 2 test. *Supplement to the Journal of the Royal Statistical Society*, 1(2):217–235, 1934.

G. U. Yule. On the application of the χ 2 method to association and contingency tables, with experimental illustrations. *Journal of the Royal Statistical Society*, 85 (1):95–104, 1922.

Appendix A

The following list containt all the keywords obtained from Corpus Nican Mopohua: Identifying lexemes (bold)

1) **Ilhui**cac (In the Sky): Qu**ilhui** (it says), Tiqu**ilhui**z (you will say it).
2) Ixpan**tzinco** (In yours presence): Yohua**tzinco** (at dawn), Mochan**tzinco** (in your dear house).
3) **Ix**quich (every thing): Quim**ix**pantilia (show something), M**ix**tzin (her dear face)
4) **Nel**tocoz (will believe) : Ni**nel**cotoz (I will be believed), I**nel**tica (Truly), **nel**li (True) and Qui**nel**tiliz (will take it)
5) **Opa**tic (healthy) : N**opa**mpa (For me)
6) **Quimahuizo** (it causes admiration) : o**quimahuizo** (what wonders), **quimahuizo**que (I cause them admiration) **quimahuizo**aya (it caused him admiration)
7) **Quimo**lhuili (he answered): **quimo**nanquili (He answered), **quimo**nanquilili (your answered), **quimo**nonochilia (communicates), **quimo**nahuatili (order it), **quimo**tlatlauhtili (begs), **quimo**lhuilia (He tells) and **quimo**ttili (He saw it).
8) **Quittaz** (will see): Ni**quittaz** (I'll see).
9) **Tepet**zintli (little hill): **tepet**l (hill)
10) **Xochitl** (Flower) : Tlazo**xochitl** (beautiful flowers).
11) **Cihuapill**i (maid), **cihuapill**é (oh my princess).
12) **Moztla** (Tomorrow): I**moztla**yoc (your morning)
13) **Ina**ntzin (his revered mother) : **ina**huac (next to someone)
14) I**nezca** (your signal): Tla**nezca**yotl(the signal of something).

15) **Itla** (something): **itla**tol (your word), Nimitzt**itla**ni (I sent you a messenger), Inet**itla**niz(Your message), **itla**nequiliztzin (his venerable will), **itla**tzin (your uncle), Inet**itla**niz (Your message), **itla**zoixiptlatzin (her venerated beloved image).

16) **Iyo**llo (your heart) : Notecu**iyo**é (Oh my Lord), Ni**iyo** (my essence), mi**iyo**tzin (your venerable breath), i**iyo**tzin (his venerable breath), Totecu**iyo** (our lord).

17) **Moch** (all) : I**moch**iuhyan (Place), **moch**iuh (happened), **moch**ihuaz (To be done).

18) Mo**nequi** (necessary): Motla**nequi**liz (for your wish)

19) **Moyollo** (your heart): **moyollo**tzin (Your dear heart)

20) Niquele**huia** (To wish): No**huia**n (everywhere)

21) No**chpocht**ziné (oh my maid) : I**chpocht**li (young woman)

22) **Notla**tol (My word) : **notla**nequiliz (My will)

23) On**can** (beech): Ni**can** (Here), itzine**can**(your beginning)

24) **Oqui**cac (he hear): **oqui**ttac (he saw)

25) **Teo**tl (God): **teo**yotl(diviniti), No**teo**cal (My tempel), **teo**pixcatlatoani (bishop), I**teo**caltzin (her venerable temple), to**teo**pixcahuan (our priest)

26) **Tlacatl**(man) : **tlacatl**é (mistress of humanity)

27) Tlat**ohuani** (the one who has a word) : Ipalnem**ohuani** ()

28) Tona**tiuh** (Sun) : Aci**tiuh** (get there), Quitca**tiuh** (I'll go see you)

29) **Ye**huatzin (excellence) : **ye**ppa (Before), **ye**tihuitz (It is coming)

30) Mot**ech** (For you): Mop**ech**tecac (lean with respect)

In the following words we have not been able to identify lexemes for various reasons, especially because of low frequencies of their occurrences.

1) Icpac (top)
2) Iquizayampa (from your place of departure)
3) Occepa (again)
4) acito (He went to get to a place)
5) Campa (place)
6) Ceme (supporter)
7) Cenquizca (entirely)

8) Cocoliztli (disease)
9) Itecpanchan (The royal house)
10) Itilma (his blanket of him)
11) Manen (I dont know)
12) Caxtillan (Spanish)
13) Nepapan (Several)
14) Nictequipachoz (afflict)
15) Nonyaz (I will go precisely)
16) Noxocoyou (my little son)
17) Omonexiti (she appeared)
18) Oniquittac (I saw it)
19) quihualtepotztocaya (he came here)
20) Zan (alone)

Chapter 2

```
################################################################
### The following measures can be computed using R
###
### 1.8 Gamma distributions

### we presented the main R codes:

############ SECTION 1.8 ###############

###### A new test for Gamma distributions package of
 Villaseñor J. and Gonzáles E. (2015) ######

##### A variance ratio estimator that provides
the asympotic null distribution is Vn*

library(fitdistrplus)
library(goft)
gamma_test( )

## The variance Vn is the ratio as a test statistic
their values are expected to be close to one

n=111
Z1=log(C)
Z1bar=sum(Z1)/n
```

```
X1bar=sum(C)/n
prod1=(C-X1bar)*(Z1-Z1bar)
sumpro1=sum(prod1)
bn1=sumpro1/n
Sigman1=X1bar*bn1
Vn1=var(C)/Sigman1


############ SECTION 1.8 AND CHAPTER 3 ###############
###### Tests of goodness of fit  gamma distribution ######
###### Tests of goodness of fit  uniform distribution ######

## logL, AIC, BIC, Chi square, AD and KS

library(readxl)
library(rriskDistributions)
res1=fit.cont( )


############ CHAPTER 3 ###############

####uniform distribution

###Empirical and theoretical CDFs by ranges

library(ggplot2)
set.seed(1)
ks.test(K,"punif",0.02,0.35)

ed <- ecdf(K)
maxdiffidx <- which.max(abs(ed(K)-punif(K,0.02,0.35)))
maxdiffat <- K[maxdiffidx]
p<-ggplot(aes(K),data=data6)+stat_ecdf()+theme_bw()
+stat_function(fun=punif,args=list(0.02,0.35))
p<-p+labs(title="ECDF and theoretical CDF")
```