Extended Entropy Maximisation and Queueing Systems with Heavy-Tailed Distributions

I.A. M. MOHAMED

PHD

2022

Extended Entropy Maximisation and Queueing Systems with Heavy-Tailed Distributions

Ismail Abdel Mageed MOHAMED

Submitted for the Degree of

Doctor of Philosophy

School of Computer Science, AI and Electronics

Faculty of Engineering and Digital Technologies

University of Bradford

2022

Abstract

Ismail Abdel Mageed Mohamed

Extended Entropy Maximisation and Queueing Systems with Heavy-tailed Distributions

Keywords: Entropy, belief functions, belief updates, uncertain reasoning, M/G/1 queue in stability phase, generalized entropies, extensive and non-extensive maximum entropy formalisms, information geometry, information length, transient queueing systems.

Numerous studies on Queueing systems, such as Internet traffic flows, have shown to be bursty, self-similar and/or long-range dependent, because of the heavy (long) tails for the various distributions of interest, including intermittent intervals and queue lengths. Other studies have addressed vacation in no-customers' queueing system or when the server fails. These patterns are important for capacity planning, performance prediction, and optimization of networks and have a negative impact on their effective functioning. Heavy-tailed distributions have been commonly used by telecommunication engineers to create workloads for simulation studies, which, regrettably, may show peculiar queueing characteristics. To cost-effectively examine

the impacts of different network patterns on heavy- tailed queues, new and reliable analytical approaches need to be developed.

It is decided to establish a brand-new analytical framework based on optimizing entropy functionals, such as those of Shannon, Rényi, Tsallis, and others that have been suggested within statistical physics and information theory, subject to suitable linear and non-linear system constraints. In both discrete and continuous time domains, new heavy tail analytic performance distributions will be developed, with a focus on those exhibiting the power law behaviour seen in many Internet scenarios.

The exposition of two major revolutionary approaches, namely the unification of information geometry and classical queueing systems and unifying information length theory with transient queueing systems. After conclusions, open problems arising from this thesis and limitations are introduced as future work.

Dedication

Through the long path, where the impossible has been proven to be possible….There was always a noble heart to always support me…To her, I dedicate my ultimate gratitude …To my wife, the hidden soldier behind my success, without whom this work would be impossible to get into birth.

In a path we start together

A light that keeps me alive

My very best friend ever

This is for sure, my wife

Acknowledgements

First but importantly, ultimately thanks to my supervisor, Dr. Qichun Zhang, for all the valuable advice, and encouragement. Through his support to for producing excellent research, and persistent efforts to disseminate knowledge and for being a real kindle to always recharge my motivation. I thank Professor Alistair Wood, my associate supervisor, for his time and insightful comments throughout our discussions. Throughout this journey, my wife's selfless love, empowerment, dedication, and support have been tremendous with no letters of any underlying language to express my gratitude.

رَبِّ إِنِّى لِمَآ أَنزَلْتَ إِلَىَّ مِنْ خَيْرٍ فَقِيرٌ

Lord! Truly, I am always in need of whatever good that You bestow on me!

The Holy Quran, Chapter 28:24

Relevant Publications

1. I.A. Mageed, and D.D. Kouvatsos, “Extended Properties of the Class of Rényi Generalized Entropies in the Discrete Time Domain”, 11-13 July 2011, IEEE *International Conference on Computer Networks and Information Technology*, Abbottabad, Pakistan, pp. 1-7, doi: 10.1109/ICCNIT.2011.6020894. Online available at https://ieeexplore.ieee.org/document/6020894

2. I. A. Mageed, and D.D. Kouvatsos, “Information Geometric Structure of Stable M/G/1 Queue Manifold and its Matrix Exponential”, Proceedings of the 35th UK Performance Engineering Workshop, 16 December 2019,School of Computing, University of Leeds, Edited by Karim Djemame, p.123-135.
   Online at: https://sites.google.com/view/ukpew2019/home

3. D.D. Kouvatsos, and I.A. Mageed (a), “Non-Extensive Maximum Entropy Formalisms and Inductive Inferences of Stable M/G/1 Queue with Heavy Tails”, in “Advanced Trends in Queueing Theory”, March 2021, Vol. 2, Vladimir Anisimov and Nikolaos Limnios (eds.), Books in ‘Mathematics and Statistics’, Sciences by ISTE & J. Wiley, London, UK.

4. D.D. Kouvatsos, and I.A. Mageed (b), “Formalismes de maximum d’entropie non extensive et inférence inductive d’une file d’attente M/G/1 stable à queues Lourdes”, “Théorie des files d’attente 2”, Mars 2021, Théorie et Practique, Sous la direction de Vladimir Anisimov et Nikolaos Limnios, Mathématique, Sciences by ISTE & J. Wiley, Londres, Royaume-Uni.

5. I.A. Mageed, and D.D. Kouvatsos, "The Impact of Information Geometry on the Analysis of the Stable M/G/1 Queue Manifold", Major extension of paper [3], 4-6 February 2021, In Proceedings of the 10th International Conference on Operations Research and Enterprise Systems - Volume 1: ICORES, ISBN 978-989-758-485-5, pages 153-160. DOI: 10.5220/0010206801530160.
6. I.A.Mageed and Q. Zhang (a), "An Introductory Survey of Entropy Applications to Information Theory, Queueing Theory, Engineering, Computer Science and Statistical Mechanics", Proceedings of 27th IEEE International Conference on Automation and Computing(ICAC), University of the West of England, Bristol, 1-3 September 2022.
7. I.A. Mageed and Q. Zhang(b), "Inductive Inferences of Z-Entropy Formalism (ZEF) Stable M/G/1 Queue with Heavy Tails", Proceedings of the 27th IEEE International Conference on Automation and Computing (ICAC), University of the West of England, Bristol, 1-3 September 2022
8. Mageed IA, Yuyang Zhou, Y, Liu, Y, and Zhang Q, " Towards a Revolutionary Info-Geometric Control Theory with Potential Applications of Fokker Planck Kolmogorov(FPK) Equation to System Control, Modelling and Simulation",In2023 28th International Conference on Automation and Computing (ICAC) 2023, 30th Aug-Sep 1. IEEE.
9. Mageed IA, Yuyang Zhou, Y, Liu, Y, and Zhang Q, "$\mathbb{Z}_{a,b}$ of the Stable Five-Dimensional $M/G/1$ Queue Manifold Formalism's Info- Geometric Structure with Potential Info-Geometric Applications to Human Computer Collaborations and Digital Twins",In2023 28th International Conference on Automation and Computing (ICAC) 2023, 30th Aug- Sep 1. IEEE.

Contents

Acronyms

ρ	Server Utilization (SU)
μ	Mean Service Rate (MSR)
λ	Mean Arrival Rate (MAR)
HQT	Heavy queue tails
QLD	Queue Length Distribution
PDF	Probability density function
CDF	Cumulative distribution function
EMC	Embedded Markov chain
MQL	Mean of queue length
i.i.d	Independent and identically distributed
ME	Maximum entropy
PME	Principle of Maximum Entropy
EME	Extensive Maximum Entropy
$\exists$	There exists a
NME	Non-extensive Maximum Entropy
CSV	Squared Coefficient of Variation
IG	Information geometry
QM	Queue Manifold

KD	Kullback's Divergence
FIM	Fisher Information Matrix
RCT	Ricci Curvature Tensor
e^A	Exponential of Matrix A
IL	Information Length

1. Introduction

Entropy as a concept is both thermodynamic and information theoretic. Indeed, entropy has influenced and inspired new innovations in communication systems, and the design of algorithms. It is valuable to note that not only in thermodynamics, but probably in all of science, entropy is a one-of-a -kind quantity.

Entropy is a concept used to quantify the level of randomness or disorder in a system. It has diverse interpretations and applications in various fields like physics, chemistry, biology, sociology, and information theory. Entropy can be defined either thermodynamically, considering system properties like temperature and pressure, or statistically, as a measure of molecular disorder. It is an extensive property, meaning it scales with the size or extent of the system, and its units are typically expressed as J/Kmol or cal/Kmol.

1.1 Motivation behind review of Entropy applications

The great scope of applications of Entropy can be difficult to analyse in depth. Therefore, for the reader to comprehend the significance of entropy, a general understanding of entropy and its applications is necessary. We also examined the justifications for entropy's inclusion in the domains of computers, engineering, queueing theory, and information theory. Entropy is an all-encompassing concept that, depending on the context, can mean several things. Additionally, a variational principle connects them fundamentally in a unique way. Entropy also exhibits some universality.

This is a reference to the idea that entropy and other metrics like compressibility or complexity often appear to be connected or overlap. This is valid, for instance, for algorithmic complexity and Lempel-Ziv complexity in information theory. The motivation for reviewing entropy applications also came from the mathematical intractability and the significance of employing entropy as a powerful tool to find better alternative solutions to many problems.

1.2 From Shannon till now, a brief

(Shannon 1948) argued that his newly invented function is unique because it is the only one that meets three fundamental axioms (Cunha et al. 2020). Generalized entropies were first introduced by (Jelinek, Tuladhar, et al., 2021; Amigó, et al., 2018; Teixeira, and L. Antunes, 2021). However, research has shown that these generalised entropies also showed new solutions subject to weak conditionalization or agreement with various axiomatizations.

Shannon's concept of average was narrower than Rényi's. There is a great scope of applicability of entropies to systems with complexity emerging from physics as well as dynamical systems stemming from suppressing the fourth Shannon-Khinchin axiom (Korbel et al. 2018). Notably, the concepts of basic space in dynamical systems theory were derived from statistical mechanics.

It is true that Shannon used a wholly probabilistic framework to prove the existence of information theory. Landauer's concept was necessary to resolve Maxwell's dilemma (Zhang et al. 2021). Furthermore, Kolmogorov's research on Bernoulli shifts was influenced by Shannon's invention. In return, symbolic dynamics has provided unanticipated benefits to information theory. Entropy and Lie ergodic theory meet these three domains at their junction, where there has been and continues to be cross-pollination.

Approximate entropy and Ergodic theory (Arbabi and Mezic 2017; Montesinos et al. 2018), directional entropy (Burget 2018,Zunino and Kulp 2017), entropy as an arrow of time (Gomes and Carneiro 2021; Donglai et al. 2018), switch entropy (Zhang and

Shang 2020), they are all proposed to address novel open problems in many scientific fields such as multi-scale analysis and synchronization, etc. For an in-depth account, including historical information, please consult the excellent review (Manish 2020) and its references.

There are many identified gaps in the existing knowledge, which was the main drives to undertake the current research in this thesis to find solutions and fill in these gaps. The work of (El-Affendi and Kouvatsos 1983) was only basic, where the Shannonian formalism of stable(non-time dependent) $M/G/1$ queue was derived. Since then, no research was done by using more generalized forms of entropies, which was a real motivation to proceed with the Rényian and the Tsallisian cases.

On another note, the lack of info-geometric analysis of queueing systems has triggered the enthusiasm to undertake this novel path of research, rather than following the classical approach to analyze queueing systems.

More importantly, in the case of Transient(or time-dependent) queueing systems, there was no existing knowledge on the information length of these time-dependent queueing systems. Therefore, it was necessary to fill in this gap, towards a novel information length theory of transient queueing systems.

It is to be noted that the contributions of this thesis have demonstrated the possible use of generalised entropy functionals in the investigation and modelling of the performance of high-speed networks with HQTs.

Because of this, the sections on engineering, computer science, statistical mechanics, queueing theory, information theory and knowledge transfer that follow provide a brief introduction of entropy applications in each of these disciplines.

1.3 Entropy applications to Information Theory

The most celebrated paper (Shannon 1948) was the corner stone of both information theory and information content, where Shannon's entropy (Cunha et al. 2020) was first introduced. First, the amount of information, or just information, is given by:

$$I = K\ ln(M) \tag{1.1}$$

In this case, *I* represent all the information included in a message (a string), and *M* represents all possible messages in a finite collection of messages that can contain the message. A constant K is used to convert from one base of logarithms to another. This is equivalent to switching between units of information (bits, trits, nuts, etc.). Therefore, by choosing the correct value for K, you can obtain any information unit using any logarithmic base. The following example shows how to find I from equation (1.1) (Cunha et al. 2020). Thus, Shannon entropy H reads as

$$H = -K \sum_i p(i) lnp(i), i = 1,2, \dots, n \tag{1.2}$$

$p(i)$ serves as the probability of symbol i. Shannon entropy H and amount of information I are linked by: $I = N.H$, where N is the number of symbols in the message.

Shannon's work was the motivation for many profound scientists to relate and develop links between thermodynamic entropy, information entropy and information. It has been revealed that information (Rex 2017) equals negentropy, which was the start to obtain the solution of Maxwell demon problem. Following Brillouin, it was announced that those irreversible logical operations (Eftimiadi and Trimigliozzi 2019), which contributed towards the establishment of the Landauer principle (Jizba and Korbel 2020). (González-Espinoza et al. 2020; Palu et al. 2020) undertook a thorough analysis to find how information and entropy relate quantitatively. The behaviour of complexity is similar to the entropy's time derivative (Theodore 2022). Moreover, the study of how thermodynamic entropy and information are related to the impact of entropy production on stochastic thermodynamic processes was carried out by (Theodore 2022; Scharfenaker and Yang 2020; Wijesuriya et al. 2021; Landi and Paternostro 2021):

$$H + I = constant = H_{max} = I_{max} \tag{1.3}$$

where H defines entropy I is information, and I_{max} serves as the maximum value of information. Consequently: “a loss of entropy is compensated by an equal information gain” (González-Espinoza et al. 2020). The maximum value of information, I_{max} is contained in a book, regardless of how much of it is known. The amount of known information I rises as we read the book, whereas the amount of unknown information, or entropy H, falls. When the book is finished, I reaches I_{max} ,which serves as the maximum value.

Shannon's entropy is a concept in information theory that measures the level of uncertainty or unpredictability in a message. It is directly related to the amount of information contained in the message. As we read a book, the known information increases, reducing the unknown information or entropy. Researchers have explored the relationship between thermodynamic entropy and information, finding that a decrease in entropy is balanced by an increase in information. This understanding forms the basis for further research discussed in chapter 4 of the thesis.

1.4 Entropy applications to Queueing Theory

Each stochastic process involves a changing probability distribution over time, which has the effect of changing the entropy or uncertainty of the probability distribution as well. This leads to a very intriguing query: Is it feasible to investigate the evolution of uncertainty over time? The birth-and-death process, which forms the basis of the typical analysis of queueing systems, requires that the mean rate at which consumers enter and exit the system be equal.

(El-Affendi and Kouvatsos 1983) used the ME formalism to analyse the $M/G/1$ and $G/M/1$ queueing systems at equilibrium and discovered the jobs' closed form expression in the $M/G/1$ queue as well as the associated service time distribution.(Guiasu 1986) used ME requirements for the known mean values to devise a probabilistic model for the underlying queue.

Following (El-Affendi and Kouvatsos 1983), the $G/M/1$ queue illustrates the queue length in a system with a general (meaning arbitrary) distribution of interarrival times and an exponential distribution of service times for each job. Whereas an $M/G/1$ queue(El-Affendi and Kouvatsos 1983), is a queue model with Markovian arrivals (modulated by a Poisson process), General distribution service times, and a single server.

To demonstrate how uncertainty fluctuates with the queue settings, we wish to create a measure of this uncertainty. Taking these factors into account, (Kapur 1986) investigated these types of fluctuations using a variety of entropy measures and came up with some intriguing findings. Research into discrete-time queues that deal with two single server queues has been done in (Kempa and Marjasz 2021). The inspiration for all these measurements came from Shannon's (Cunha et al. 2020) fundamental entropy measure. Several of these measures were implemented because of (Jelinek and Tuladhar, et al., 2021; Kapur 1986; Chakraborti et al., 2021; Rapisarda et al., 2019; Tahmasebi, 2020; Deppoman, 2020; Wang et al., 2020; Toomaj and Di

Crescenzo, 2020), among other factors. These measurements can help in understanding how uncertainty behaves in the various queueing system states, as demonstrated by (Singh et al. 2021).

The concept of maximum entropy, sometimes known as PME, is a potential tool for analysing complex queuing models in a variety of contexts. Combining the principle of sufficient reason, PMEs can be used to determine equilibrium probabilities with respect to the underlying distribution's moments constraints. An exact queue length distribution may be derived, as in $M/M/1$ system model.

A considerable number of publications have helped shape the idea of using PME for the analysis of different schemes of queues. The measurement of uncertainty is the idea of maximum entropy, which Shannon first stated in the context of information theory (Cunha et al., 2020), and which Jaynes later expanded (Jizba and Korbel 2020). Cross-entropy maximisation and PME in system modelling are axiomatized by Shore and Johnson (Ghosh 2020). (Cantor 1984) considered a multi-server queue that was in a steady state and gave information theoretic analysis based on PME for a queue that was running concurrently. The maximisation of entropy was used by (Kouvatsos 1985; Jain et al. 2021) to discuss general queuing networks. By engaging PME, (Kouvatsos 1986;Guiasu 1986), the maximum entropy flow in networks is computed in Gabriel (2017). The applicability of maximal entropy to the moment's problem was highlighted by (Zhang et al. 2020).

The analysis of vacation queueing models in various frameworks was covered by (Panta et al.2021). Notably, a novel closed form expression for the maximized entropic solution for a stable $G/G/1$ queue was obtained by (Kouvatsos 1988), which has revealed the impact of moment constraints on the derivations. The ideal entropy analysis was derived by (Huang et al. 2018). Additionally, (Chauhan 2018) employed PME to analyse the queue size distribution for the unstable $M^X/(G1, G2)/1$ model with the Bernoulli vacation schedule.Numerous works that addressed PME-based queuing models were constrained to dealing with certain queuing scenarios, and the results were frequently acquired in implicit form.

Briefly, the principle of maximum entropy (PME) is a tool used to analyze complex queuing models in various contexts. It can be used to determine equilibrium probabilities with respect to the underlying distribution's moments constraints, and an exact queue length distribution may be derived. Many publications have explored the use of PME for the analysis of different queuing schemes, and it has been applied to various queuing models, including the $M/G/1$ and $G/M/1$ queueing systems, as well as general queuing networks.

1.5 Applications of Entropy to Engineering

Even if the applicability of entropy as a concept has been extended over many fields (Kapur 1993), there are still further areas where it could be used. There are several packages of entropy in commercial engineering. For example, research problems have been studied subjectively by proposing a suitable entropy-primarily totally based technique, such as the dispatching problem (Marvizadeh 2013) of a fabric dealing with gadget in terms of Automatic Guided Vehicles (AGV) in a discrete component production gadget. The dispatching problem manner allocating AGVs to facilitate shifting requests to assure the green component glide withinside the manufacturing facility. Based on Kullback-Leibler directed divergence, an entropy-primarily based total aid allocation algorithm that effectively recollect the result of ability actions at the load stability of the manufacturing facility earlier than the allocation of resources had been employed (Marvizadeh 2013).

The suggested algorithms are appropriate to implement and attempt to reduce the burden inside the manufacturing unit even as pleasing the flow requests generated with the aid of using the manufacturing unit paintings centres. Additionally, (Marvizadeh 2013) investigates rating and choice very well primarily based totally on the implied overall performance measure. Combining PME and Kullback-Leibler directed divergence concepts to provide an algorithm composed of two stages for this problem. The proposed technique can no longer anticipate any priori distribution assumed with the aid of using Bayesian methods, and in the end, it affords a rating of structures primarily based totally on their located overall performance measures. Moreover, an entropy-primarily totally based criterion for evaluating alternatives. The comparison (Marvizadeh 2013) is primarily based totally on directed divergence among alternatives' cumulative possibility distributions between being introduced. To identify the causal connections in complex structures by considering the internal composition alignment of the temporal structure, a novel information-theoretic measure known as the coupling entropy was employed (Zhao and Shang 2015). To determine the degree of uncertainty in the coupling between two time series, an asymmetric association measure based on permutations is used. In the examination of Hénon maps, where various noises are introduced to assess its accuracy and sensitivity, it is discovered that the coupling entropy is effective. Characterizing neural membrane mutual coupling with the help of entropy-based iterative learning identification has been devised (Tang et al. 2020). (Melin et al. 2020) used the concept of design complexity to examine the concept of entropy's substantial significance to engineering design theory and methodology.

In summary, Entropy is a concept that has been applied in various fields, including engineering. In manufacturing systems, entropy-based algorithms have been developed to allocate resources and optimize the flow of components. Additionally, entropy has been used to identify causal connections in complex structures, such as neural membrane mutual coupling, and to examine the significance of entropy in engineering design theory and methodology.

1.6 Applications of Entropy to Computer Science

The conceptualization of Information is a mathematical discipline (Goodfellow et al.2016) that worried about transmitting information through a loud channel. In standard, this may be hired to calculate entropy, which quantifies the statistics in an occasion and a random variable. Concepts are worried with information compression, which is called supply coding. In addition, it offers that transmitting and storing it in a manner this is strong to errors, that's called mistakes correction or channel coding (Avand and Moradi 2021). Additionally, entropy is employed as the premise for strategies that include characteristic selection, constructing choice trees, and, in a wider standard scope, becoming type models (Brownlee 2020). Following (Goodfellow et al.2016), low information(unsurprising) would contribute to a **high probability event**, whereas high information(unsurprising) is a descriptor of a **low probability event**.

The primary instinct supporting information theory is learning the occurrence of a likely event is less informative than learning the occurrence of an unlikely event. An entropy-based image segmentation approach (Barbieri et al. 2009) was introduced and an application to Google Earth’s extracted colour images was provided**.**

It is the available entropy variations that kindled the motivation (Asokan and Anitha 2019) to reveal that Tsallis entropy has been found preferable for thresholding and clustering.

Investigating the strength of cryptographic keys is in fact a continuous challenge for both academic and industrial practise. The difficulties in generating and evaluating the available entropy for cryptographic needs are discussed in (Ershadi et al. 2019). Furthermore, (Ershadi et al. 2019) encouraged the development of new spectrum estimate algorithms by considering conventional entropy estimation in cryptography applications.

Based on the theoretical analysis and experimental results, it is found that the Exploration Entropy (Xin et al. 2020) contains more information in comparison to the existing analytical methods used to analyse and manage the training process of Reinforcement Learning (RL), which illustrates the strength of this new technique.

To summarize, information theory is a mathematical discipline that deals with transmitting information through a noisy channel. Entropy measures the information's amount in an event or a random variable. It has applications in image segmentation, cryptography, and reinforcement learning, where it is used to improve the efficiency and accuracy of algorithms.

1.7 Applications of Entropy to Statistical Mechanics

Boltzmann's pioneering work (Merriam 2021) is credited with investigating the characteristics of gas bodies and seeing them as systems made up of many molecules. Properties like total volume, total molecule count, and total energy are examples of macro-states. The position and speed of individual molecules are examples of the qualities that define micro-states. We propose the existence of gas body’s molecules in separate states to make the explanation more straightforward.

Boltzmann applied the "principle of indifference", which was applied by Boltzmann, where it is assumed that all micro-states occur with equal probability. The formula for Boltzmann’s distribution can now be used to investigate various properties of gas bodies.

It is noted (Jizba and Korbel 2020) that Boltzmann's logic can be extended to situations unrelated to statistical mechanics and reinterpreted using information theory. Statistical mechanics "may be stemming from statistical inference," he warned. To measure how uncertain, we are about the system, Jaynes (Shore and Johnson 1980) proposed employing Shannonian entropy to replace Boltzmann's thermodynamic entropy. It is important to remember that PME asserts that, given known descriptive statistics, the probability distribution with the highest entropy best embodies the state of the art in statistical inference.

In brief, Boltzmann's pioneering work investigated the characteristics of gas bodies and saw them as systems made up of many molecules. He proposed the existence of gas body's molecules in separate states to explain the position and speed of individual molecules, which define micro-states. Boltzmann's principle of indifference assumes that all micro-states occur with equal probability, and his formula for Boltzmann's distribution can be used to investigate various properties of gas bodies. Additionally, Boltzmann's logic can be extended to situations unrelated to statistical mechanics and reinterpreted using information theory, which can measure the uncertainty of a system.

1.8 Entropy Applications to Knowledge Transfer (KT)

Many people hold the opinion that knowledge transfer (KT) is an essential step in knowledge management. Examples of these crucial types of KT include intergenerational knowledge transfer (IGT) (Cyr and Choo 2010; Leistner 2010; Nonaka and Takeuchi 1995; Bratianu 2011; Lefter et al.2011). The fact that the KT process involves intricate interconnections across the three basic categories of knowledge—rational, emotional, and spiritual knowledge—should be given more consideration (Bratianu, 2018a; Bratianu and Bejinaru, 2019, 2020). A presentation of knowledge entropy (KE) as a concept and a description of how it may be used to measure complexity based on the connection with the thermodynamic phenomena discussed above, of information management and to explain knowledge transfer. We define KE as follows, based on the formula for information entropy introduced by Shannon in 1948:

$$KE = -C \sum_i p_i \, ln p_i \tag{1.4}$$

Here, the measurement is calibrated to a certain scale and environment using an arbitrary positive constant called C. We can suppose that $p_1, \dots, p_n$ it depicts the knowledge's distribution within any organisation by hypothesizing that all employees own the knowledge or being sources of the knowledge. You can consider that distribution using both space and time. Information entropy reaches its greatest value and knowledge dynamics are in equilibrium when all these probabilities are equal. But since each person is unique and has unique experiences, emotions, and spiritual levels, such a circumstance is not feasible. Since there is not a knowledge management measure that can quantify absolute knowledge for everyone, relative values for knowledge that are specified in relation to a particular level of knowledge can be utilised instead. When evaluating the effectiveness of KT within programmes for training or for information-sharing activities inside organisations or communities of practise, the concept of KE can be quite helpful (Wenger 1998). Managers can decide on their tactics for raising knowledge entropy, which promotes creativity, by measuring the distribution of knowledge inside a particular department or organisation. Researchers can also demonstrate methods for fostering intellectual capital and building intelligent organisations by assessing knowledge entropy (Bratianu 2018; Bratianu 2013).

According to Bratianu (2019), KE depicts the likely distribution of knowledge within a certain organisation at a particular period. Even while we think of organisational knowledge as being like a field, it rests with specific individuals, yielding an individual knowledge's distribution for a specified time. The innovation process and the fundamental competences that contribute to competitive advantage are impacted by the dynamic nature of this distribution and its variations. From the perspective of mathematical modelling, knowledge entropy and information entropy are comparable, but from a semantic perspective, they are completely unrelated.

The modern trend of research only investigates KE without conducting empirical research, which led to its limitation. The knowledge probability distribution function has a somewhat ambiguous interpretation when viewed mathematically, and there are some useful techniques for gathering important information and calculating the knowledge entropy indicator in each situation and at a specific moment. Additional study should focus on creating workable techniques for calculating knowledge entropy and knowledge distribution probability sets. To further understand the relationship between KE as a concept and a certain organization's performance, more empirical research is also required.

The topic of measuring recognition knowledge—particularly categorization knowledge—and how it changes was discussed in (Hou 2018). Three premises served as the foundation for this discussion. The equation suggested to characterise the impact of knowledge on uncertainty was used to generate two formulas for evaluating the levels of knowledge for recognition in two scenarios. In addition, by looking into how ignorance evolves in the face of uncertainty, the idea of knowledge entropy was established, and its formula was given. We looked at its resemblance to Boltzmann's entropy and how it differed from Shannon's entropy. Based on a mathematical examination, proof was found to back up the following conclusions:

- Learning results in a reduction in knowledge entropy.
- The distinctiveness of the people's rating order increases with decreasing knowledge entropy.
- A group's collective knowledge level is not always equal to the sum of its members' individual knowledge levels.
- If a person's desire for knowledge never increases, their knowledge entropy will never rise.

The two most frequent knowledge transfer (KT) techniques used in organisations and organisational networks are personalization and codification (Sudhindra et al 2017). (Sudhindra et al 2017) have suggested a theoretical model of KT to analyse organisations' (KT) processes in terms of the method for exchanging tacit knowledge (i.e., gained knowledge without living the actual experience) and the associated information content. The concepts of tacitness and information content, as well as their impact on the choice of KT methods, have been described using Shannon's entropy, an idea from information theory. Notably, (Sudhindra et al. 2017) has aided with:

- Using information content to anticipate the choice of the KT mechanism.
- Development of an intuitive explanation for the tacit-explicit continuum as well as a tacitness expression.
- The creation of a KT theoretical model that can be put into practice to predict which KT mechanisms will be used in real-world situations, as well as the characterization of product variety in terms of information content.
- Figure 1.1 depicts the KT model (Sudhindra et al. 2017), which combines personalisation and codification techniques. As can be seen, the KT process is determined by two critical constructs: tacitness and the volume of information content. When people and businesses operate across larger geographic areas, individualised interactions become more expensive, and codification techniques take over as the standard KT. However, three key challenges may limit the implementation of personalisation. Individualised interactions become more expensive as people and businesses operate across larger geographic areas and codification tactics become the standard KT. However, there are three key barriers that can restrict the use of personalization.

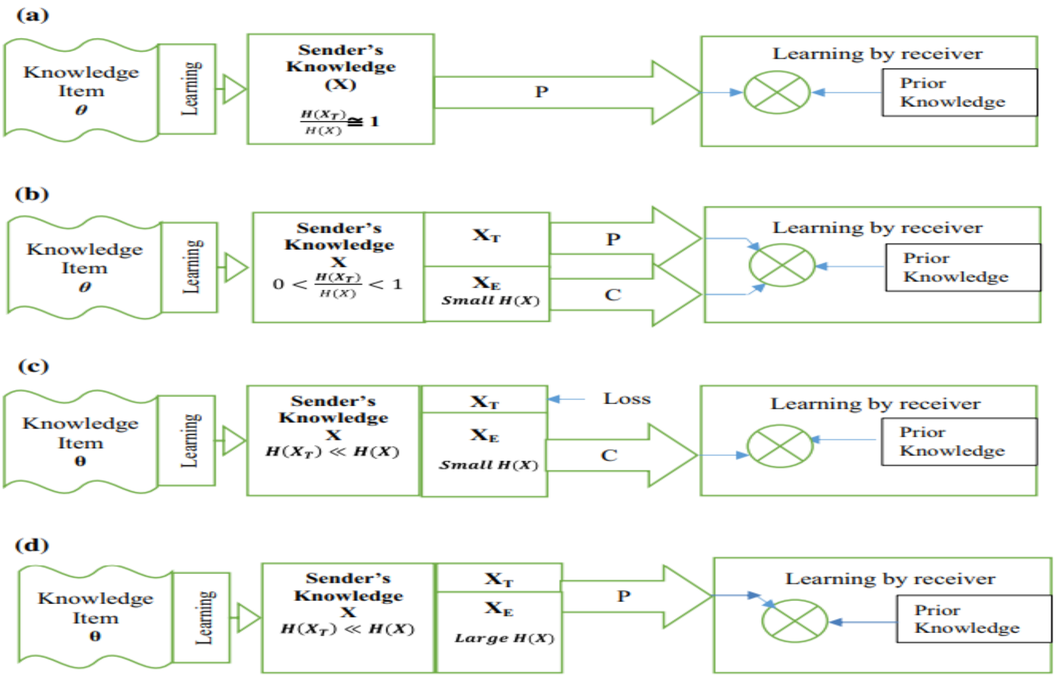


Figure 1.1. Schematic depiction of the KT model (Sudhindra et al. 2017)

So, section 1.9 discusses the importance of knowledge transfer (KT) in knowledge management, including intergenerational knowledge transfer (IGT). The concept of knowledge entropy (KE) is introduced to measure the complexity of knowledge distribution within an organization, which can be useful for evaluating the effectiveness of KT programs. The KT model (Sudhindra et al. 2017) is also presented, which combines personalization and codification techniques and is determined by the constructs of tacitness and information content.

The aim of this research is to develop a cutting-edge approach to unify uncertain reasoning theory with information theory. More potentially, to reveal the information theoretic impact on queueing systems and to open new grounds to contemporary queueing theory through information geometry and information length theory.

To this end, the main objectives of the current research include:

- To investigate RGEs properties in the discrete case through an analytic approach.
- To maximize Rényi entropy under some constraints to reveal the closed form non-extensive formalism of the stable $M/G/1$ and to reveal the unknown impact of non-extensivity to the performance of the underlying queue,
- To trigger and inspire how a stable queue, namely $M/G/1$ is impacted by information geometry?
- To set up a first-time ever mathematical investigation to uncover the information length of a time-dependent (Transient) queue, namely, $M/M/\infty$.

1.9 Contributions of the thesis

The current thesis has added several major contributions to the corpus of existing knowledge, which are listed as follows:

- Advancing the class of Rényian Generalized Entropies Extended Properties and Finding PV-updates in the Discrete Time Domain are obtained[1]. Over the past few years, the literature has mostly focused on the continuous-time domain when describing the properties of the family of Rényi's generalized entropies (RGEs) that are categorized as being information theoretic, particularly as probabilistic procedures for inductive inference.

[1]This provides the solution to the open problem investigating the extension of the class of Rényian Generalized Entropies' properties and Finding PV-updates in the Discrete Time Domain (Mageed and Kouvatsos 2011).

- An original extension of these properties to the discrete-time domain, a generalization of an analytical result on the interpretation of maximum entropy (ME) formalism as a consistency requirement, and a methodology for determining probability vector updates (PV-updates) based on previous information-theoretic findings on minimum cross entropy. Most importantly, this work would significantly add to the knowledge since it opens new doors to the discretization of the RGEs properties rather than being confined to examining these properties from a continuous domain perspective.
- The provision of novel comprehensive unification of information theory and queueing systems, which is proven to be mathematically credible by employing the four consistency axioms of inductive inference[2].

  In mathematical terms, a new knowledge regarding the information theoretic influence of the non-extensive parameter for analyzing stable queueing systems has been established.
- Technically speaking, the upper and lower bounds of the data information length of a transient $M/M/\infty$ queuing system are derived to present a novel connection between Information Length Theory (ILT) and Transient Queueing Systems (TQSs). In this context, it is shown that the upper and lower bounds acquired are both (n,t)-dependent, with *n* = 0,1, 2,..... If t serves as the time-dependent server usage of the transient $M/M/\infty$ queuing system, then the latter derived upper and lower bounds ($UB(n,t), LB(n,t))$ respectively) are both dependent on n.

[2]This is a breakthrough which unifies Information theory with Queueing Theory, by showing the non-extensive information theoretic impact on the overall performance of stable queueing systems (Kouvatsos and Mageed 2021a).

A typical numerical experiment is also proposed to demonstrate the influence of time on the behaviour of the developed $UB(n,t)$ and $LB(n,t)$ for various values of n. As a result, two new state probabilities—the Rényian and Tsallisian formalisms for the stable M/G/1 queue—have been produced because of a demonstrated significant influence. To make the newly discovered solutions precise, new underlying families of underlying Rényian and Tsallisian service probability density functions(PDFs) as well as cumulative distribution functions were derived. More intriguingly, it is demonstrated that the impact of information theory also applies to the newly created squared coefficients of variation in the Rényi and Tsallis situations.

- More importantly, we have demonstrated the plausibility of our derived solutions by demonstrating how the consistency axioms were used to reason about them. Mathematically speaking, we demonstrated that three axioms are satisfied and that the non-extensivity influence only led to the defiance of one axiom. This is a giant step towards revealing the hidden information-theoretic impact on the performance of stable queueing systems.
- A new discovery of Info-geometric Queueing Theory (IGQT) is devised[3]. This revolutionary approach reveals how queueing theory can be combined with diverse areas of mathematics, such as differential geometry, information theory, matrix theory, and information geometry (IG).

---

[3] As far as the author's knowledge allows, the influential info-geometric role in analysing stable queue manifolds that is revealed by the author of this thesis for the first time ever in literature(Mageed and Kouvatsos 2019; Mageed and Kouvatsos 2021).

More crucially, the application of info-geometric approaches to queueing theory offers a potent method for examining queue stability and enabling the use of cutting-edge IG techniques to fundamentally alter traditional queueing theory.

- The determination of the information length of the $M/M/\infty$ transient queue, which contributes to the unification of Information Length Theory (ILT) and Transient Queueing Systems, we make a significant advancement towards the quantification of the number of statistical variations corresponding to a specified temporal range for transient $M/M/\infty$ transient queueing systems (TQSs).

1.10 The Structure of the Thesis

Chapter 2 presents the supporting information for the contributions of this thesis. We begin by introducing Rényi entropy and uncertain reasoning. The second section explains ME, along with its derivation for discrete ME distributions. The Performance Analysis of Transient Queueing Systems and a summary of ME solutions for queueing system performance distributions are presented. This chapter ends with a succinct, detailed explanation of information geometry and theoretical context for information length. Advancements to the class of Rényi's Generalized Entropies Extended Properties and Finding PV-updates in the Discrete Time Domain are characterized in Chapter 3. In Chapter 4, it is demonstrated that a heavy-tailed stable $M/G/1$ queue exists using non-extensive maximum entropy formalisms and inductive reasoning. The impact of info-geometric analysis on the stable $M/G/1$ queue manifold is determined in Chapter 5. In Chapter 6, both lower and upper bounds of information length of the transient $M/M/\infty$ transient queueing system are determined. Chapter 7 draws conclusions combined with limitations and emerging open problems from this study are discussed. Following chapter 7, we have Appendices of the detailed proofs. Chapter 7 is followed by appendices of the detailed proofs.

2. Review of literature

The underlying information for the contributions of this thesis is presented within the current chapter. We start by overviewing Rényi entropy and uncertain reasoning. The second section then introduces the Maximum Entropy (ME) principle and includes its derivation for discrete ME distributions. The third section contains a summary of ME solutions for queueing system performance distributions, focusing on both GGeo and truncated GGeo ($\mathrm{GGeo_T}$) ME solutions, as well as a summary of Performance Analysis of Transient Queueing Systems. In the fourth section, a short review on stable queueing systems with two real-life applications of the transient $M/M/\infty$ queueing system are provided. In the fifth section, a concise, in-depth description of information geometry is provided. A conceptual background for information length is introduced in the sixth section. Existing research gaps, combined with aims and objectives are highlighted in section seven. Finally, section eight provides a summary of the chapter.

2.1 Rényi Entropy and Uncertain Reasoning

The development of novel measures for quantifying complexity in the time-frequency plane and signal information was motivated by Rényi's generalized entropies(RGEs). RGEs strongly resemble the concept of complexity that is utilised when visually inspecting time-frequency images(Flandrin et al. 2018; Saulig et al. 2017).

RGEs (Flandrin et al. 2018; Saulig et al. 2017)are defined to be the family of entropy measures, $\mathrm{H_{q,R}(p)}$

$$\mathrm{H_{q,R}(p)} = \frac{\mathrm{c}}{1-\mathrm{q}} \ln \sum_{n=0}^{\infty} (p(n))^{q} \quad , q \neq 1 \tag{2.1}$$

For any constant c$> 0, p(n)$ to denote the steady state probability at state n.

These measures are excellent candidates for time frequency analysis because they have a number of additional intriguing and beneficial characteristics, including accounting, cross-component, and transformation invariances. In (Saulig et al. 2017), a thorough investigation of the characteristics combined with an illustration of applicability of RGEs is presented, with a focus on the mathematical underpinnings of quadratic time-frequency representations. It was determined that there are signals for the Wigner distribution for which the measurements are not clearly specified (Saulig et al. 2017).

(Zitnick 2003) provided an expanded entropy and mutual information estimator for Rényi's definitions of entropy, which include Shannon's concepts as special instances. To effectively maximise Rényi's quadratic entropy in response to a defined set of linear equations restrictions, a measure that is a member of the family of Rényi's entropies was proposed (Zitnick 2003) by forbidding the probability estimates to be in the range from 0 to 1.

It was hypothesized (Anderson et al. 2021) that coherent-state inputs reduce the Rényian outcome entropy. Notably, (Sason 2018) suggested a theorem to generalize the Tunstall codes (Tunstall 1968) utilizing the Rényi's entropy, which enhances the importance of this determination and builds a formula to reduce redundancy. Furthermore, the repercussions of two various meanings of shared info to the generalization of capability and price distortion work were analysed (Sason 2018) and prolonged the credibility of Tunstall's theorem (Tunstall 1968) to countable alphabet.

The range of Rényi's inference procedures in distinct opportunity was discovered (Hawes 2007) to have frontiers of minimax at some point as well as the restricted centre of the mass inference procedure CM_∞ at the various other points. Formulas for determining minimax as well as maximin could be discovered in (Hawes 2007),which rank over those of ME for inferring ideas worth, which are logical varieties when the agent's understanding is on its own revealed simply in logical varieties.

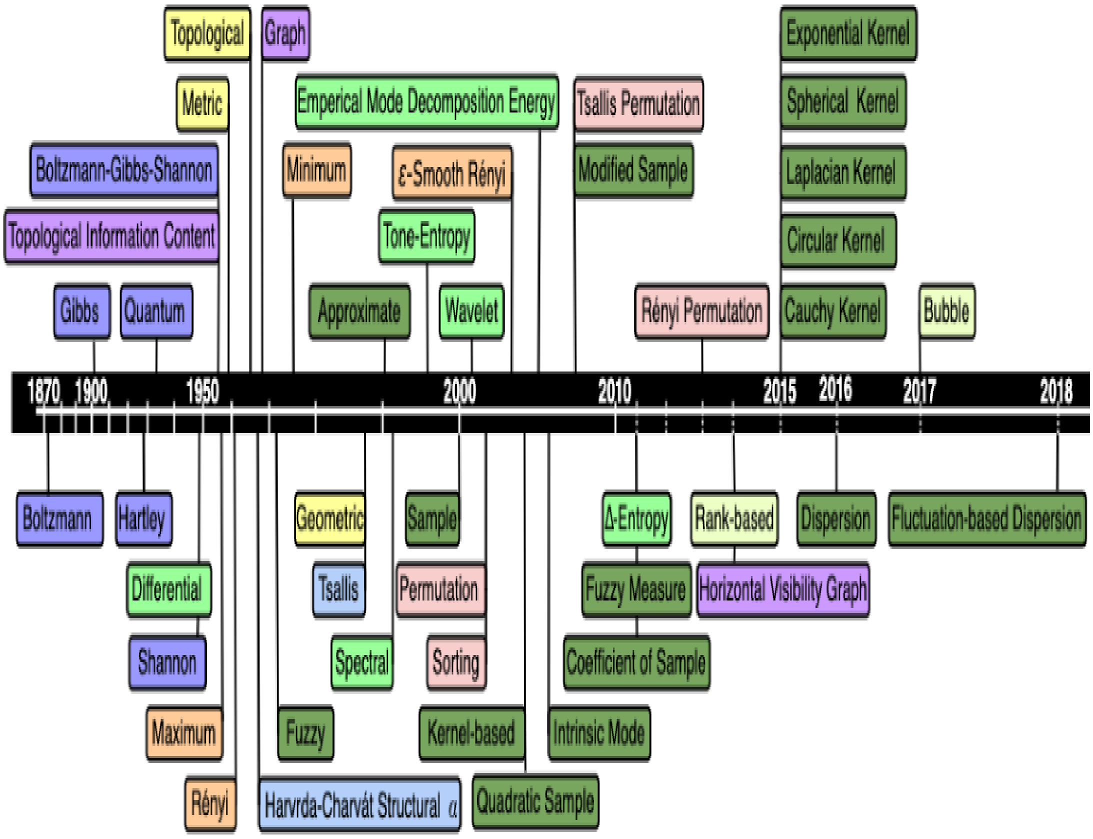


Figure 2.1: Timeline of entropy, (c.f., Ribeiro, et al.2021)

In the circumstance of the nonparametric issue of estimating Rényi's entropy as well as shared info (MI), based upon a finite example attracted coming from an unidentified, constant circulation over R^d (the d- fold up Cartesian item of genuine varieties R), certainly there are lots of requests that utilize such estimators. Each entropy estimator as well as shared info estimators have been utilized for subspace evaluation through (Van Hulle 2008) as well as picture enrolment through (Uffink 1995). Furthermore, a course of estimators for the Shannon's as well as Rényi's info of multi-dimensional possibility thickness were made on a proposal in (Nanda and Choudhury 2021).

Notably, (Kumar and Choudhury 2011) have also utilised their newly proposed L_{α}^{β} measure to demonstrate a coding theorem for a discrete noiseless channel.

Probability theorists typically refix the issue through presenting some ‘natural' presumptions based upon the ideas of equi-probabilities, self-reliance and so on. If these presumptions are actually ‘close' to truth, the possibility the service is actually frequently ideal, in which situations utilizing an idea features method is of no utilization. A design standing for quantified ideas, based upon supposed idea features, was presented of (Dempster 1997) as well as opened the method to the derivation of novel results, some which are currently checked listed below. Details that outline computational problems like appraisal located bodies, quick Möbius change as well as approximate techniques can be discovered (Paris 1994). Details that the Shafer's design (Paris and Vencovska 1992; Jiroušek and Shenoy 2018) offered the structure for the advancement of a brand-new mathematical concept of proof, based upon using idea features. This design, as opposed to those based upon possibility techniques of inference (Jelinek, and Tuladhar et al. 2021;Tunstall 1968) was utilized in various self-controls (nonetheless, it is based upon various interpretations (Jiroušek and Shenoy 2018; Van Hulle 2008), like

1.A possibility design (Van Hulle 2008) based upon the mathematical concept of tips (Nanda and Chowdhury (2021).

2.A transferable idea design (Smets et al.1991)

The complication in between these different interpretations discusses very most mistakes experienced in the literary works, where writers evaluate the rooting concepts of Shafer's (Dempster 1997;Smets et al.1991) design.

The Belief (Bel) function (Paris 1994) is defined by assuming that it obeys the probability axioms. Fix a finite propositional language L and call the set of sentences for this language SL. In this case,

For all $\theta, \phi \epsilon SL$, we define Bel: SL → [0,1] as a probability function on SL and it should satisfy the following:

1.If $\vdash (\theta \leftrightarrow \emptyset)$, then Bel$(\theta) = Bel(\emptyset)$

2.If $\vdash \theta,$ then Bel$(\theta) = 1, and\ Bel(\neg\theta) = 0$

3.I$f \vdash (\theta \wedge \emptyset)$ is false, then Bel $(\theta \vee \emptyset) = \mathrm{Bel}(\theta) + Bel(\emptyset)$

Let us define the set $V^L(K)$ (Paris 1994)., $V^L(K) = \{x^{\rightarrow} \in R^J / x^{\rightarrow} A_{K = b_k^{\rightarrow}}, x^{\rightarrow} \geq 0\}$

2.2 Conceptualization of ME

As a statistical inference technique, Jaynes developed the ME principle (Jaynes 1957; Jaynes 1978). More interestingly, "frequentist" and "Bayesian" are the two core techniques in statistical inference.

The latter, (Cox 2006) typically incorporates a universal probabilistic concept. Subjective (personalistic) and objective Bayesian methods are subsets of the Bayesian method. Probabilities are viewed as representations of (rational) people's levels of individual belief in subjective Bayesianism. Probabilities, on the other hand, are viewed as representations of a state of knowledge in objective Bayesianism that is unaffected by an individual's personality.

According to the frequentist perspective, in a random experiment, probabilities are seen as observable and verifiable frequencies (Jaynes 2003).

According to the ME principle, this inference technique is objective Bayesian, because prior knowledge is considered and probability assignments are provided independently of people's personalities and emotions, as well as independent of experiments.

Notably, ME suggests, based on prior knowledge, inferring the probability distribution of a random quantity, which is frequently assumed to take the form of the random quantity's moments. Prior moment and normalization (Jaynes 1957) data are frequently satisfied by a wide variety of distributions. So, which of these distributions should be chosen to arrive at the best or most logical conclusion? The data-driven distribution. The "unbiased" (or "least biased") distribution is defined in this context as the distribution that is "maximally non-committal to unknown information".

Now, we arrive at the Shannonian entropy, H (known to be the distribution's entropy), which is read as

$$H(\dots p_{-2}, p_{-1}, p_0, p_{1,} p_2 \dots) = -\sum_{n=-\infty}^{\infty} p_n ln p_n \, , n = \cdots, -2, -1, 0, 1, 2 \dots \qquad (2.2)$$

where the stationary occurrences or state probabilities are represented by the p_n's, p_n',$n \in \mathbb{Z}$.

Shannon's entropy is a valid information measure because it meets several postulates, including those stated in (Csiszár 2008), with references providing the relevant proofs.

The prior information must be accurate enough to allow us to determine whether the least biased inference meets it. Moments and boundaries are two examples of such prior knowledge. "The first moment of the random value (RV) is presumably less than 0.6," for example, is inadmissible information.

Given previous moment data, the ME principle (Shore and Johnson 1980) generates only one distribution inference. See (Jaynes 1978; Fang et al. 1997) for a more detailed account.

Consistent differences between the inferred distribution and the observed probability point to a poorly constrained ME issue formulation in which all erroneous information has been associated to the constraints (Jaynes 1968). There is a wide range of ME applications to numerous scientific disciplines. See (Shore and Johnson 1980).

2.2.1 Distributions of discrete ME

Given that this thesis uses the ME principle in that context, discrete ME distributions are produced below. Consider a RV, N, when it comes to modeling any quantity in the discrete domain of any underlying system or process, using the derivation in (Jaynes 1957), then the probabilities p_n= $P(N = n)$, n =. . . , −2, −1,0,1,2 … are unknown and need to be found. N maps integers a sample space, S, for example, S = $\mathbb{Z}$. Assume that the only prior knowledge of normalization and repenting N by its moments, $E[f_i(n)]$,i = 1,2,3 … m .

Probability averages or sample moments can be used as the moments of the distribution inferences, $E[f_i(n)]$.It is to be noted that Bayesian and likelihood criteria, as well as hypothesis testing, have been used to establish the validity of employing sample moments (Jaynes 1978). Maximizing Shannon's entropy (c.f., (2.2)) subject to the constraints

$$\sum_{n=-\infty}^{\infty} f_i(n)p_n = E[f_i(n)] \tag{2.3}$$

Normalization is determined by:

$$\sum_{n=-\infty}^{\infty} p_n = 1.0 \tag{2.4}$$

is a constrained nonlinear optimization problem that can be solved with the help of a Lagrange multiplier. The Lagrange function $L(p), p = \left(i \dots p_{-2}, p_{-1}, p_0, p_{1,}p_2 \dots\right)$,can be written in the form:

$$L(p) = -\sum_{n=-\infty}^{\infty} p_n ln p_n + \sum_{i=0}^{m} \beta_i (\sum_{n=-\infty}^{\infty} f_i(n) p_n - E[f_i(n)]) \qquad (2.5)$$

Here, β_0 serves as the Lagrangian multiplier linked to normalization, whereas m moment constraints are related to the Lagrangian multipliers β_i. Thus, for all n, $f_0(n)$ = 1. By setting $\partial L/(\partial p_n)$ = 0, n =. . . , −2, −1,0,1,2, …, the state probabilities, $p_n's$, at which entropy is maximised are determined. Hence, ME solution is devised by:

$$p_n = e^{-\sum_{i=0}^{m} \beta_i (\sum_{n=-\infty}^{\infty} f_i(n) p_n)}, n = . . ., -2, -1,0,1,2, \ldots m \qquad (2.6)$$

Denoting the normalization constant as $(1/Z)$, equation (2.6) can be written as

$$p_n = \frac{1}{Z} e^{-\sum_{i=0}^{m} \beta_i (\sum_{n=-\infty}^{\infty} f_i(n) p_n)}, n = . . ., -2, -1,0,1,2, \ldots m \qquad (2.7)$$

If Z is given by the normalising constant's inverse,

$$Z = \sum_{n=-\infty}^{\infty} e^{-\sum_{i=0}^{m} \beta_i (\sum_{n=-\infty}^{\infty} f_i(n) p_n)}, n = . . ., -2, -1,0,1,2, \ldots m \qquad (2.8)$$

The following partial derivative can be used to obtain $\beta_i (\text{c.f.}, \text{Equ.}(2.6)) i = 1,2,3 \ldots m$.

$$-\frac{\partial \beta_0}{\partial \beta_i} = E[f_i(n)],\ i = 1,2,3 \ldots m \qquad (2.9)$$

where $\beta_0 = lnZ$.

In mathematical terms, we can introduce the Lagrangian coefficients, $x = e^{-\beta_i}$.
Additionally, the following general product from a discrete ME distribution is produced by the latter replacement in (2.7).

$$p_n = \frac{1}{Z} \prod_{i=1}^{m} x_i^{f_i(n)}, n = . . ., -2, -1,0,1,2, \ldots m \qquad (2.10)$$

where $Z = \sum_{n=-\infty}^{\infty} (\prod_{i=1}^{m} x_i^{f_i(n)})$.

Consider the scenario where the initial moment of N, $E(N)$, is known. This moment is represented as an information constraint by:

$$\sum_{n=0}^{\infty} np_n = E(N), n = 0,1,2 \ldots \quad (2.11)$$

The normalization is read as

$$\sum_{n=0}^{\infty} p_n = 1.0 \quad (2.12)$$

Consequently, the discrete ME distribution is thus obtained as follows:

$$p_n = x_1^n, n = 1,2, \ldots m \quad (2.13)$$

$x_1 = \lambda/\mu$, with both λ being the mean arrival rate and μ as the mean service rate of the underlying queueing system.

When the normalising condition (2.12) is used, the formula for (2.13) can be changed to

$$p_n = (1 - x_1)\, x_1^n, n = 1,2, \ldots m \quad (2.14)$$

It represents the well-known modified geometric. Additionally, when the first moment formula is used, we can rewrite (2.13) as

$$p_n = \left(\frac{1}{1+E(N)}\right)\left(\frac{E(N)}{1+E(N)}\right)^n \quad (2.15)$$

2.3 ME Queueing System Performance Distributions Solutions

The length of the queue as well as ME solution have identified residence and waiting times as queueing system performance variables for inferring distributions. The QLD of the $M/M/1$ queue can be seen in equation (2.13)(Beneš 1965; Cantor et al. 1986).

Following Shore (1982), ME solutions matching all the performance distributions of the $M/G/1$ queue included the MQL, p_0 with more other appropriate constraints. In the instance of the $M/M/1$ queue, several of these more recent ME methods seem to be precise.

Additionally, (Lopez-Herrero 2002) has investigated ME solutions corresponding to $M/G/1$ retrial queue including both first and second ordinary moments. For different scenarios for the establishment of ME solutions for different categories of queueing systems, see (Walstra 1985; Cantor et al. 1986).

The ME solution to the $QLD's\ of\ M/G/c, G/M/c, and\ G/G/c$ queues included the MQL and set of state probabilities $\{p_0,\ p_1,\ \ldots\ p_{c-1}\ \}$ (Wu and Chan 1989). It was discovered that the ME solution offered an accurate inference of the QLD of the $G/M/c$ queue. The MQL, mean buffer length, and P($all\ c\ servers\ busy$) were employed in a ME solution to precisely describe the QLD of the M/M/c queue in Arizona et al. (1991). The odds of having at least j tasks of class I in service and approximate marginal $MQLs$ were used in ME solutions(Kouvatsos and Tabet-Aouel 1994) to approximate the $QLDs$ of multiple class $G/G/c$ queues following the pre-emptive resume(PR)scheduling discipline.

(Wang et al. 2002; Ke and Lin 2008; Yang et al. 2011) have employed both MQL and server utilisation in ME solutions through approximating the marginal $QLD's$ of the N-policy $M/G/1$ queue with removable server.

2.3.1 ME Analytic Solutions for the Typical Infinite-Capacity Queues

ME solutions have been developed by (Shore 1982; El-Affendi and Kouvatsos 1983; Guiasu 1986) to address QLD of the standard $M/G/1$ queue. The mean service time rate, μ and the squared coefficient of variation(SCV), C_s^2 represent a fundamental set of well-known queueing parameters in these ME solutions. The generalised geometric (GGeo) ME QLD was developed for a stable $M/G/1$ queueing system in (El-Affendi and Kouvatsos 1983). Along with the MQL and normalisation prior information requirements, the GGeo expressly includes the queue stability. When a queue's average effective arrival rate matches its average effective departure rate, as measured by

$$\lambda = \mu(1 - p_0) \tag{2.16}$$

where p_0 is the lower boundary state probability and λ serves as the mean-arrival rate. When $\lambda > \mu$, instability results.

As a previous moment information constraint, the queue stability requirement can be added as follows[4]

$$\sum_{n=0}^{\infty} h(n)p_n = p_0 = 1 - \frac{\lambda}{\mu}, h(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \ldots \end{cases} \tag{2.17}$$

Using p_0 and first moment constraints, ME formalism (2.10) provides GGeo expression.

$$p_n = \begin{cases} \frac{1}{Z}\left(\frac{1}{y}\right), & n = 0 \\ \frac{1}{Z}x^n, & n = 1,2,3 \ldots \end{cases} \tag{2.18}$$

Putting $(1/yz) = p_0$ in (2.18) would transform (2.18)to be in the form

$$p_n = \begin{cases} p_0 & , n = 0 \\ p_0 y x^n, & n = 1,2,3 \ldots \end{cases} \tag{2.19}$$

Note that p_0 is equated to $1/(Zy)$. Thus, the result in the re-parameterised GGeo (Kouvatsos 1988) is:

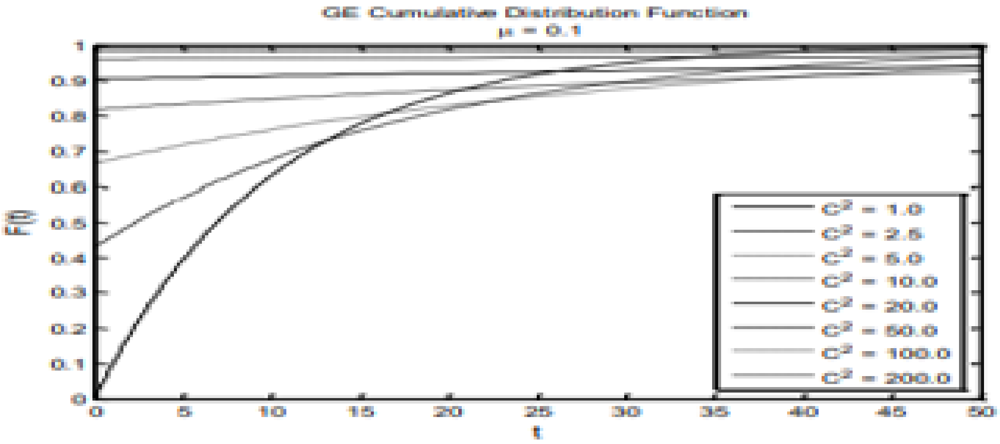
---

[4]Since they are both comparable prior information constraints that result in the same ME inference, the server utilisation constraint, '$1 - p_0$' ,can be used in place of the 'p_0' prior information constraint.

$$p_n = \begin{cases} p_0 & , \ n = 0 \\ (\frac{(1-p_0)^2}{E(N)-(1-p_0)})(\frac{E(N)-(1-p_0)}{E(N)})^n, & n = 1,2,3 \ldots \end{cases} \tag{2.20}$$

The GE CDF that makes p_n of (2.20) exact (El-Affendi and Kouvatsos 1983) is determined by:

$$F_t = 1 - \frac{2}{1+C_s^2} e^{-\frac{2}{1+C_s^2}\mu t}, t \geq 0, \mu > 1 \tag{2.21}$$



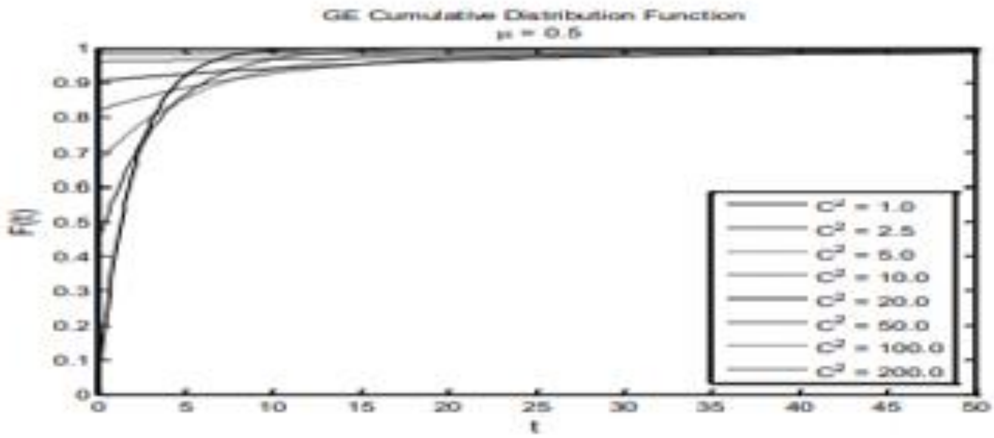


Figure. 2.2 : (GE) CDF corresponding profiles to mean rates of μ = 0.1 and μ = 0.5, respectively, from top to bottom, C_s^2= [1, 200].

For C_s^2 >1.0, the GE CDF has been thoroughly examined (Kouvatsos 1988; Kouvatsos 1994)[5].The $GE/GE/1$ queue, which has matching first two moments for C_s^2, C_a^2 > 1.0 was studied by (Kouvatsos 1988; Kouvatsos and Tabet-Aouel 1994).

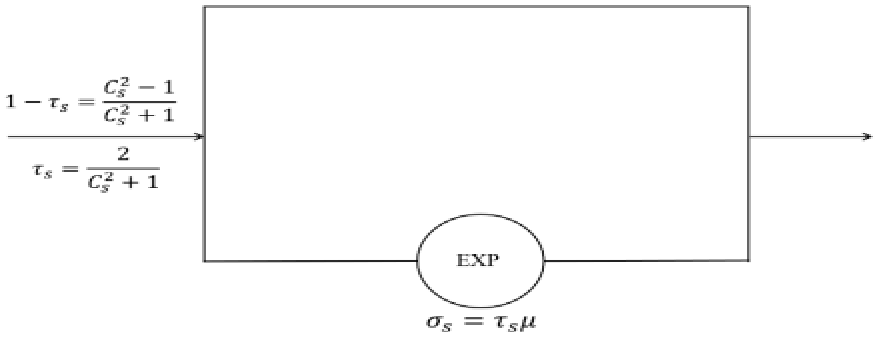


Figure 2.3: The GE-type service time distribution with parameters (for $1/\mu$, C_s^2 >1).

2.3.2 ME closed form analytic representations for Typical Finite-Capacity Queues

There has also been analysis using the ME principle for the $QLD's$ of several stable queues with finite capacity(Kouvatsos 1986b; Kouvatsos 1986a). The stability condition for the $G/G/1/K$ queue is written as

$$\lambda(1 - p_K^K) = \mu(1 - p_0^K) \tag{2.22}$$

p_0^K and p_K^K serve as the temporal fractions at which the underlying queueing system with finite-capacity is empty(full), respectively.

Additionally, we have λ and μ. Prior moment information constraints[6] can be used to describe the boundary state probabilities, p_0^K and p_K^K, as follows:

---

[5]The GE distribution's interpretations and comparative performance bounds when $C_s^2 < 1.0$ are provided in (Kouvatsos 1988; Kouvatsos 1994).

[6]The server utilisation constraint, $(1 - p_0^K)$, is equal to the p_0^K prior information constraint in the infinite capacity situation.

$$\sum_{n=0}^{K} u_K'(n) p_n^K = p_0^K = \frac{\mu-\lambda(1-p_K^K)}{\mu}, \qquad u_K'(n) = \begin{cases} 1, & n = 0 \\ 0, & n = 1,2 \dots K \end{cases} \tag{2.23}$$

and

$$\sum_{n=0}^{K} f_K(n) p_n^K = p_K^K = \ p_0 = \frac{\lambda(1+p_0^K)-\mu}{\mu}, f_K(n) = \begin{cases} 1, & n = K \\ 0, & n = 1,2 \dots K-1 \end{cases} \tag{2.24}$$

Respectively.

According to (Kouvatsos 1986a; Kouvatsos1986b), the $\mathrm{GGeo_T}$ is determined by

$$p_n^K = \begin{cases} p_0^K, & n = 0 \\ p_0^K \ y_K (x_K)^n, & n = 1,2,3 \dots K-1 \\ p_0^K \ z_K y_K (x_K)^K, & n = K \end{cases} \tag{2.25}$$

These asymptotic approximations are used in future research to roughly analyse complex queueing systems using the ME principle; see, for example (Kouvatsos and Almond 1988; Kouvatsos and Denazis 1993; Kouvatsos and Awan et al. 2003).

Considering the $\mathrm{GGeo_T}$ distribution (2.25) of their values given by, $E[N_K]$, $p_0^K$and p_K^K, if we equalize the expressions of suitable moments, $\sum_{n=0}^{K} f_K(n) p_n^K$, n = 0, we obtain

$$\frac{(1-p_0^K-p_K^K)(1-x_K)}{(1-(x_K)^{K-1})} \left(\frac{1-(x_K)^{K+1}}{(1-x_K)^2} + \frac{(K+1)(x_K)^K}{(1-x_K)} - K(x_K)^{K-1} \right) + Kp_K^K - E[N_K] = 0, \tag{2.26}$$

$$y_K = \frac{(1-p_0^K-p_K^K)(1-x_K)}{p_0^K x_K (1-(x_K)^{K-1})} \tag{2.27}$$

and

$$z_K = \frac{p_K^K}{p_0^K (x_K)^K y_K} \tag{2.28}$$

Based on (Kouvatsos 1986b; Kouvatsos 1986a), the expression for p_0^K and the stability requirement can then be used to obtain an asymptotic approximation for the parameter z_K (2.22) (Kouvatsos 1986b).

2.3.3 Performance Analysis of Transient queueing Systems

Queuing systems frequently use time-dependent parameter adjustments. The theory of time-dependent queueing system analysis is credited to (Kolmogorov 1931). The usefulness of time-dependent queueing systems has been raised numerous times in a wide range of scientific fields. As a result, such an analysis is difficult because Little's law and other well-known steady-state queueing system relations must be reconstructed (Bertsimas and Mourtzinou 1997).

Performance evaluation systems are commonly developed in response to real-world problems. As a result, many publications describe an assessment approach as well as how it is used to address a real-world problem. A variety of techniques were used to assess no spatial dimension time-dependent performance of single-stage queuing systems (Alfa 1979; Chung and Min 2014; Alnowibet and Perros 2006; Van de Coevering 1995; Tarabia 2000).

There are several approaches of performance evaluation within the literature. The Chapman-Kolmogorov equations refer to a system of differential equations (DEs) that interpret the dynamic behaviour of a Markovian queueing system (CKEs). The $M(t)/M(t)/c$ system's dynamic behaviour is described by the following DEs:

$$p_0'(t) = \mu(t).p_1(t) - \lambda(t).p_0(t), \quad n = 0$$

$$p_n'(t) = (n+1)\mu(t).p_{n+1}(t) + \lambda(t).p_{n-1}(t) - \big(\lambda(t) + n\mu(t)\big).p_n(t), \quad 1 \le n < c$$

$$p_n'(t) = c\mu(t).p_{n+1}(t) + \lambda(t).p_{n-1}(t) - \big(\lambda(t) + c\mu(t)\big).p_n(t), \quad n \ge c \qquad (2.29)$$

The exact solutions for equations (2.28) only exist in some special cases, for example, when $c \to \infty$. It is to be noted that the numerical solution of these DEs could be undertaken either by Runge Kutta or Euler techniques. It has been suggested that we could approximate systems with infinite waiting rooms(i.e., $c \to \infty$) by employing a system with a sufficiently large waiting room that is finite.

By restructuring (2.29), we might be able to get the k-th moment differential equation (MDE):

$$E[L^s(t)]' = \sum_{n=0}^{\infty} n\, p_n'(t) = \lambda(t) - c\mu(t) + \mu(t). \sum_{n=0}^{c-1}(c-n)p_n(t) \tag{2.30}$$

$$Var[L^s(t)]' = \sum_{n=0}^{\infty}(n - E[L^s(t)])^2\, p_n'(t)$$

$$= \lambda(t) + c\mu(t) - \mu(t). \sum_{n=0}^{c-1}(2E[L^s(t)] + 1 - 2n)(c-n)p_n(t) \tag{2.31}$$

where $E[L^s(t)], Var[L^s(t)]$ serve as the first moment and the variance of $M(t)/M(t)/c$ system respectively.

2.3.3.1 Modelling with piecewise constant parameters-based techniques

These techniques are divided into the following Piecewise stationary models (PSM) descriptors:

(i)**PSM with independent periods**

The approaches vary when it comes to determining the analysed interval length l and the input parameters that correspond to performance computations. The advantage of these techniques is based on their low complexity in computations, specifically in the case of existing exact solutions for steady state for the underlying system configuration.

(ii) **PSM with linked periods**

(Stolletz 2008a;Stolletz 2008b) extended the stationary backlog -carryover approxi-mation(SBC) to analyse the $M(t)/G(t)/1$ system. For a more detailed explanation (Selinka, et al. 2006; Stolletz 2013).

(iii) Piecewise transient models(PTM)

These techniques are transient models- based (TMB) that are employed in the analysis of consecutive intervals whose input parameters are constant. See (Choudhury, et al. 1997; Upton and Tripathi 1982) for more information.

More fundamentally, modified system characteristics-based techniques can be categorized as:

(i) Modified overload approach (MOL)

The number of jobs in an infinite -server system is interpreted by MOL techniques through leveraging the DE's well-known solution. There are four network models in this framework: queueing, fluid, diffusion, and simulation. To elaborate, asymptotic queueing models have traditionally been used. For more information, see (Feldman, et al. 2008; Ingolfsson, et al. 2007; Yom-Tov and Mandelbaum 2014).

(ii) Modified job characteristics (MJC)

In the MJC category, the fluid approximation (FA), the pointwise stationary fluid flow approximation (PSFFA), and the diffusion approximation (DIFF) replace the discrete task with the continuum. Different methods are used to explain probability theory. Uniform acceleration (UA), a fourth method, modifies job arrival and service rates. The idea behind fluid approximation is to replace randomly arriving discrete jobs with a deterministic continuum. This continuum can be thought of as a time-varying fluid flowing into a reservoir. Deterministic reservoir outflow approximates service processes. The reservoir's fluid level approximates the number of jobs in the system; for a more detailed explanation, see (Aguir et al. 2004; Hampshire et al. 2009; Worthington and Wall 1999).

To increase traffic density. Figure 2.4 depicts such transformations used in the analysis of the time-dependent queueing systems. As a result of the Coordinate Transformation Technique (CTT), the performance at the end of an interval is used as an initial condition in subsequent interval fluid approximations CTT. Dependencies between successive approximations are included in CTT, as presented for time-dependent queues (Mauro and Pompigna 2020).

The time-varying traffic intensity shape, on the other hand, is limited to rectangular peaks and off-peak periods with zero traffic intensity. SBC and CTT consider system performance's temporal behaviour, including dependencies between consecutive intervals. It can also be used to provide the performance evaluation of overloaded systems for a limited time (Kimber and Hollis 1978; Stolletz 2008a).

The CCT describes the overload situation using a model based in part on the deterministic fluid approximation, which provides an accurate performance approximation for the congestion period.

The gained additional relevance of the FA approach is justified by the fact that FA is an integral part of other analytic techniques, for example CTT, and within PSFFA, which is explained in the following paragraph.

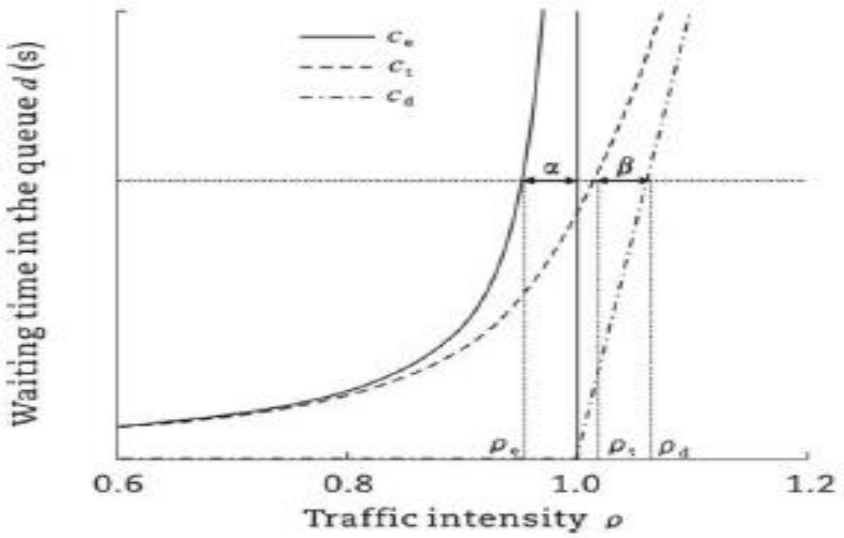


Figure 2.4: coordinate transformation method with respect to ρ

PSFFA theory provides an interesting combination between steady-state queueing formalism and deterministic fluid approximation and integrates stochastics. For a more detailed survey on PSFFA, see (Agnew 1976, Chen, et al. 2013; Cosmetatos 1976; Xu, et al. 2014).

In DIFF, the interactable discrete stochastic process $L^s(t)$ is replaced mathematically with a Brownian motion(a continuous stochastic process $\chi(t)$). The stochastic process $\chi(t)$ is prescribed for any non-empty system by the diffusion equation (known as the Kolmogorov or Fokker-Planck) equation:

$$\frac{\partial f(x,t)}{\partial t} = \frac{a(t)}{2}\frac{\partial^2 f(x,t)}{\partial x^2} - b(t)\frac{\partial f(x,t)}{\partial x} \qquad (2.31)$$

Eq. (2.31), which uses x as a continuous estimate of the number of jobs in the system and $f(x,t)$ as a time-dependent probability density function. The partial differential equation (2.31) has straightforward solutions depending on $a(t), b(t),$ and the boundary conditions; otherwise, it needs to be solved numerically (Clarke 1956;

Czachórski, et al. 2010; Czachórski, et al. 2009; Di Crescenzo and Nobile 1995;Duda 1986; Massey and Pender 2013).

It is to be noted that the arrival and service rates are raised concurrently while maintaining a constant ratio by the uniform acceleration (UA) technique. UA could be thought of as steady-state analysis's non-stationary counterpart (Massey 2002). The PSA, the fluid approximation, and the diffusion approximation are all rigorously justified by the UA results. These findings imply that the PSA performs well for underloaded queues (Flick and Liao 2010), and that the fluid approximation performs well as an approximation for overloaded queues(Mandelbaum and Massey 1995). These results also support the central tenet of the CTT.

Time-dependent queueing systems have several uses in service, aviation and land traffic, and IT systems, among other fields (Schwartz, et al. 2016). The currently in use assessment techniques frequently rely on discrete-time methods, stationary models, or fluid approximations. Notably, certain evaluation techniques are restricted to a particular field of application. For example, the PSFFA is mostly utilised for truck handling facilities, whereas the analysis of road traffic systems mostly uses CTT. Only the theoretical work of (Parlar 1984) addresses service rates' optimization, which is a possible area for further study. Another unexplored area is the time-dependent choice of whether to provide waiting areas, which (Hampshire, et al. 2009) introduce in the context of a call centre. In conclusion, this review demonstrates that time-dependent queues have a wide range of applications.

The following section provides a short review on queueing systems with two real-life applications of the transient(time-depenent) $M/M/\infty$ queueing system.

2.4. Short review on stable queueing systems with two real-life applications of the transient $M/M/\infty$ queueing system

2.4.1 Basic queues

A basic queueing system is a service system where customers arrive and require service from servers. Customers may join a queue if all servers are busy, and the order in which customers are served is determined by the queue discipline. Queueing models are mathematical descriptions of these systems, making assumptions about arrival and service processes, number of servers, and queue organization. These models help calculate performance measures to design or improve service systems.

In queueing theory, utilization is a measure of productivity and is calculated as the average number of busy servers divided by the total number of servers. Higher utilization levels lead to longer wait times, with delays increasing at an increasing rate as utilization increases. System size and variability also play a role, with larger systems experiencing shorter delays and higher variability leading to longer delays at any given utilization level. These principles have implications for capacity planning and evaluating service systems (Green 2003).

The Poisson process is a commonly used model for arrivals in queueing systems. It assumes that customers arrive one at a time, the probability of arrival is independent of when other customers arrived, and the probability of arrival at a given time is independent of the time. This model is often applicable in various contexts, such as emergency rooms and customer service call centers, and can be tested for goodness of fit using statistical measures (Green et al 2005).

In the context of steady (non-time dependent) queues, namely these queues with non-time dependent probability density function. For example, the M/M/s or Erlang delay model is a commonly used queueing model in service systems. It assumes a single queue with unlimited waiting room that serves s identical servers. In this model, customer arrivals follow a Poisson process with a constant rate, and the service duration, such as the length of stay or provider time for a patient, is assumed to have an exponential distribution. These assumptions are referred to as Markovian, hence the use of the two "M's" in the model notation (Green and Nguyen 2001).

The M/M/s system assumes an exponential distribution for service time, where the coefficient of variation (CV) is equal to one. Even if the actual CV of service time is slightly different from one, the M/M/s model can still provide reasonable estimates of delay. However, if the CV is significantly different from one, the M/M/s model may either underestimate or overestimate actual delays. In such cases, when the arrival process follows a Poisson distribution and there is only one server, the average delay can be calculated using the M/G/1 system formula, another well-known none-time dependent queue (Green et al 2005).

More fundamentally, the transient(time-dependent) queue follows a time-dependent probabilistic distribution, as an example, we have $\boldsymbol{M}/\boldsymbol{M}/\infty$ queue.

2.4.2 $M/M/\infty$ queueing system and the computation of the common average time for unsaturated site visitors flows beneath double-parking situations.

Double parking (DP) violations of industrial trucks whilst they load and dump at transport places with inadequate curb side area may have huge poor effect on site visitors. Motivated by the need to examine such effect on urban streets, (Gao and Ozbay 2016) make use of parking violation facts for New York City in conjunction with discipline facts accrued the usage of video recording and adopts a complete modelling technique that mixes to be had facts with varieties of fashions. Another implied approach was the micro-simulation version (Gao and Ozbay 2016) advanced and calibrated to examine personal and mixed outcomes of diverse explanatory variables. The research undertaken (Gao and Ozbay 2016) employed a macroscopic $M/M/\infty$ queueing version and micro-simulation for estimating common tour time with inside the presence of double-parking activities.

Under uncongested site visitors` situations without downstream blocking off, making use of the $M/M/\infty$ queueing version produced an amazing match with the sphere facts. Overall, the $M/M/\infty$ queueing version is a powerful technique to compute the common average time for unsaturated site visitors flows beneath double-parking situations. Micro-simulation is a greater effective device than the $M/M/\infty$ queueing version for comparing such congested situations and may be used to study person and mixed outcomes of diverse explanatory variables.

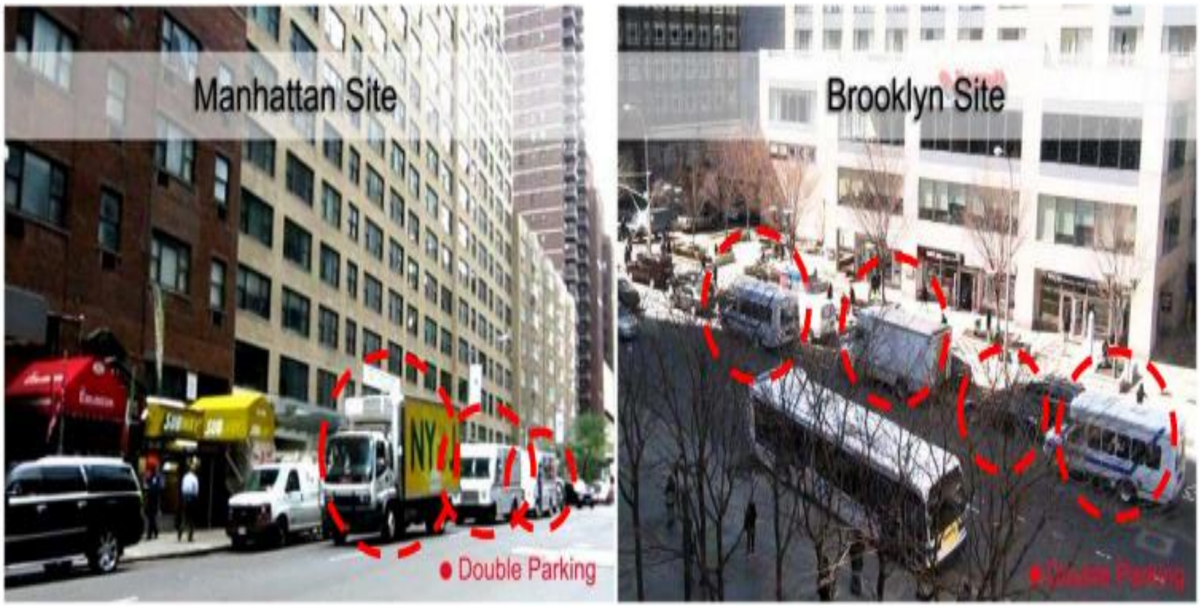


Figure 2.5. How DP occurs in the sites under investigation(Gao and Ozbay 2016)

2.4.3. Application of $M/M/\infty$ birth–death process to quantitively interpret the Wavelet Dynamics in Atrial Fibrillation and Phase singularity.

It has been hypothesized that the determined range of PS or wavelets in AF might be ruled with the aid of using a not unusual set of renewal rate constants λ_f (for PS or wavelet formation) and λ_d (PS or wavelet destruction), with steady-state population dynamics modelled as an M/M/∞ birth–death manner. It has been demonstrated (Dharmaprani et al. 2021):

(1) that λ_f and λ_d may be blended in a Markov $M/M/\infty$ manner to as it should be a version of the determined common range and population distribution of PS and wavelets in all structures at unique scales of mapping; and

(2) that slowing of the constants denoting rates, namely λ_f and λ_d is related to slower mixing rates of the $M/M/\infty$ birth–death matrix, presenting an interpretation of spontaneous AF termination.

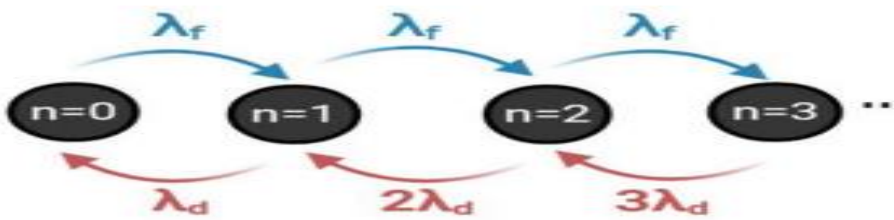


Figure 2.6. The birth-death process of a transient $M/M/\infty$ queueing system (Dharmaprani et al.2021)

It is worth mentioning that λ_f and λ_d (PS rates of formation and destruction respectively) are related by the following steady-state equation of the $M/M/\infty$ birth–death process (Kleinrock 1976)

$$N = \frac{\lambda_f}{\lambda_d} \tag{2.32}$$

provided that N serves as the average number of PS and wavelets. Additionally, the PS and wavelet population distribution is characterized by the steady state probability p_n (Kleinrock 1976) of getting a wavelet population or a phase singularity with size n is determined by:

$$p_n = \frac{(\frac{\lambda_f}{\lambda_d})^n e^{-\frac{\lambda_f}{\lambda_d}}}{n!} \tag{2.33}$$

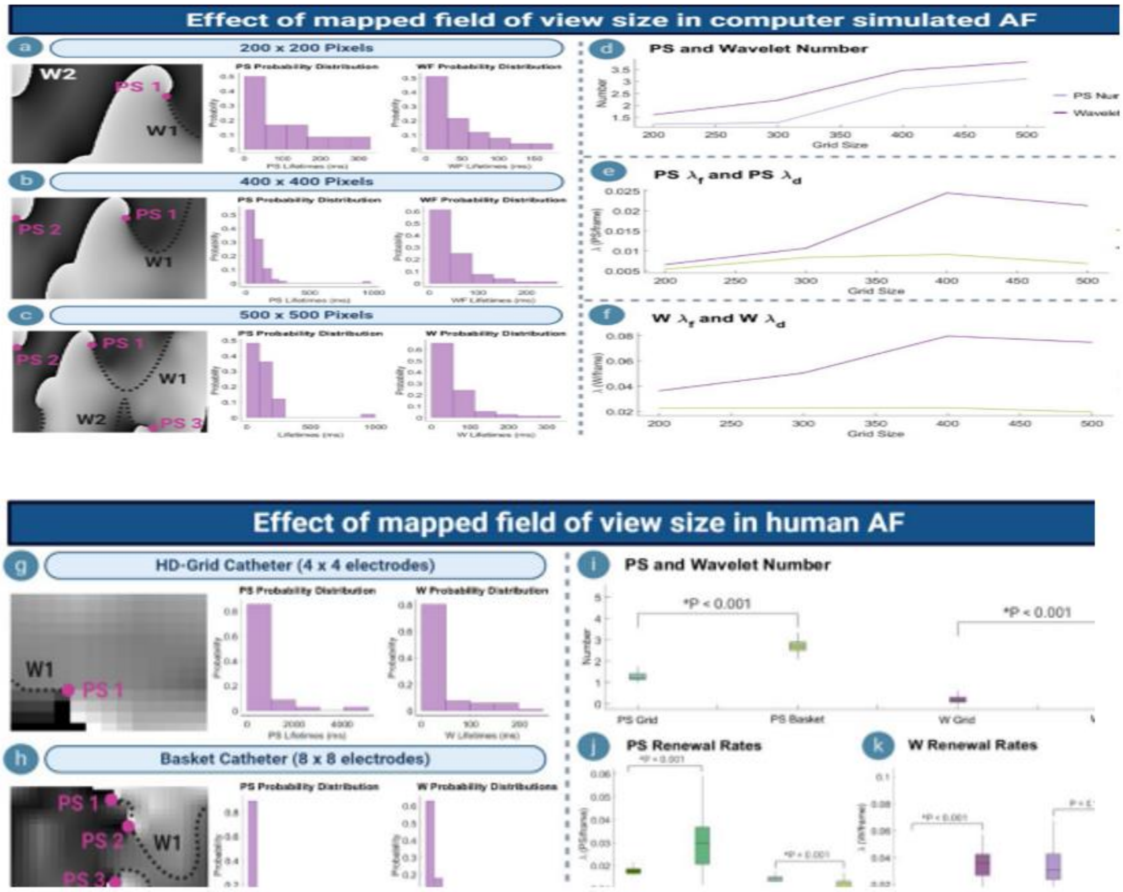


Figure 2.7. How a mapped field impacts view size (Dharmaprani et al.2021)

It has been conjectured by (Dharmaprani et al.2021) that both equations (2.32) and (2.33) provide a strong characterization of the overall dynamics of PS and wavelet population.

$M/M/\infty$ birth–death techniques offer a singular quantitative representational framework to conceptualize and recognize PS and wavelet population dynamics in AF. This conceptual paradigm (Quah et al., 2020) has been proven to be used in all types of AF studies, at a lot of one-of-a-kind scales and densities of mapping, offering possibilities for scientific application.

2.5 Information Geometry

Information geometry (IG) (Amari, 1985) is based on the application of non-Euclidean geometry approaches to stochastic processes and probability theories. IG rethinks a probability distribution family in terms of a statistical manifold (SM). Additionally, IG revolutionises the way we describe probability density functions by studying their corresponding SMs to enable the provision of the geometric metric. This overall novel approach is significantly crucial as it visualizes SM as a coordinate system. A manifold is a topological finite-dimensional Cartesian space, $\mathbb{R}^n$, where an infinite-dimensional manifold exists (Baez 2021). $\mathbb{R}^n$ is, as can be seen, a topological space. Surprisingly, the description of SMs is intuitively supported by IG. As a result, the visualisation of the derived geometric figures demonstrates the significance of IG (Nielsen 2020).

Amari (1985) defined information geometry as the application of non-Euclidean geometry approaches to stochastic processes and probability theories. In terms of a statistical manifold, IG rethinks a probability distribution family (SM). IG has also been used to investigate SMs, where the geometric metric provides a novel description of the probability density function, which is critical in SM and can be visualised as a coordinate system. A manifold is a finite-dimensional topological Cartesian space, in which an infinite-dimensional manifold exists (Baez 2021). As can be seen, $\mathbb{R}^n$ is a topological space. Surprisingly, IG intuitively supports the description of SMs. As a result, visualising the derived geometric figures demonstrates the importance of IG (Nielsen 2020).

Figure 2.8 depicts $\widehat{\theta}$ (the parameter inference) of a model from data and looks at it to represent a decision-making algorithm: If we have a category of models $M = \{m_\theta\}_{\theta\epsilon\{\theta_1,\theta_2,..,\theta_n\}}$ that best fits the data, then what are the parameters $\{\theta_1, \theta_2, .., \theta_n\}$ to make this holds for the probability density function of the geometric manifold distribution under investigation? The differential-geometric manifold structure M of IG can be used to create decision rules. (Amari 1985) investigated the exponential distribution families, whereas (Dodson 1999) studied and revealed the geometric structures of some special exponential distributions.

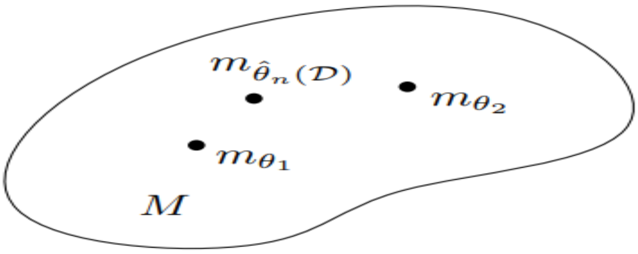


Figure 2.8: SM Parametrization (c.f., Nielsen, 2020)

The (IME) is a square matrix that is analogous to the ordinary exponential function and is used to solve linear differential equation systems. Furthermore, the matrix exponential is important in Lie groups (Hall 2015).

The real motivation for us taking this path of research was based on IG of a stable M/D/1 queue and introducing a geometric structure on the set of M/D/1 queues by utilising the properties of queue length paths (Nakagawa 2002), which is, to the best of our knowledge, the only research paper in the literature. Moreover, the geometric approach was used to study the invariance and equivariance of figures in a coordinate-free approach (Kondor and Trivedi 2018).

Ricci curvature is used to measure the deviation of the Riemannian metric (RM) from the standard Euclidean metric (EM) (Nielsen 2020). Furthermore, scalar curvature is used to calculate the difference in volume between a geodesic ball and a Euclidean ball of the same radius.

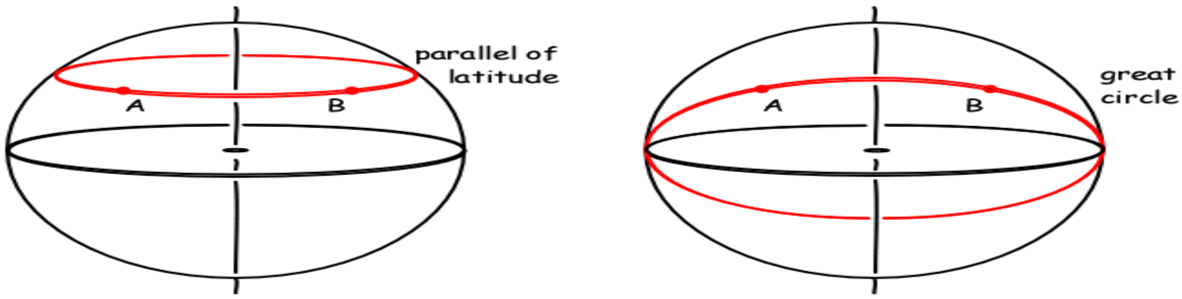


Figure 2.9: On curved surfaces, geometric geodesics (Norton 2020).

The RM tensor of a statistical model's parameter manifold is defined by the Fisher information metric. We can possibly utilize it to compute gap of information between measurements.

Kullback's Divergence, or KD (Regli and Silva 2018), is a relatively simple objective to optimize. However, because KD considers the log-likelihood ratio p/q, it tends to penalize the region more where $p > q$ — that is, overestimating the true posterior is penalized more than underestimating it for any given region. The derived approximation tends to undercover regions of low probability in the target model (Turner & Sahani, 2011), while focusing on several modes based on the constraints. (Amari, 2012) has developed power EP (Minka 2005) and the black-box alpha divergence (Hernandez-Lobato et al., 2016) in the context of variational inference. The

Rényi divergence (Rényi et al., 1961; Van Erven and Harremos, 2014) will be the primary focus.

$$D_R^{\gamma}(p||q) = \frac{1}{(\gamma-1)} ln\left(\sum_{n=0}^{\infty}\big(p(n)\big)^{\gamma}\left(q(n)\right)^{1-\gamma}\right) \tag{2.34}$$

Variational Inference VI by Rényi is investigated in (Li and Turner 2016). More advancements on theory of traditional VI on complex models were determined by (Depeweg et al., 2016).Outliers in the training data also cause problems for KD (Ghosh et al. 2017). See (Basu et al. 1998; Ghosh and Basu 2016) for a more detailed account on different categories of divergence measures.

The AB-divergence was introduced and investigated by (Cichocki et al. 2011):

$$D_{s,AB}^{\gamma,\eta}(p||q) = \frac{1}{\eta(\eta+\gamma)} ln\left(\sum_{n=0}^{\infty}\big(p(n)\big)^{\gamma+\eta}\right) + \frac{1}{\gamma(\eta+\gamma)} ln\left(\sum_{n=0}^{\infty}\big(q(n)\big)^{\gamma+\eta}\right) - \frac{1}{\gamma\eta} ln\left(\sum_{n=0}^{\infty}\big(p(n)\big)^{\gamma}\left(q(n)\right)^{\eta}\right) \tag{2.35}$$

for $(\gamma, \eta)\epsilon\mathbb{R}^2$ such that $\gamma \neq 0, \eta \neq 0\ and\ \ \gamma + \eta \neq 0$.

The authors (Cichocki et al. 2011) have presented a novel (dis)similarity measure, namely $\boldsymbol{D_{s,AB}^{\gamma,\eta}(p||q)}$ (c.f., (2.35)) , where it has been illustrated (Cichocki et al. 2011) that $D_{s,AB}^{\gamma,\eta}(p||q)$ is potentially robust. More intriguingly, the recent extension of classical linear metric learning methods has taken two distinct paths: deep metric learning (Cilingir et al. 2020) methods for learning data embedding using neural networks, and Bregman divergence learning approaches for learning more general divergence measures such as divergences over distributions.

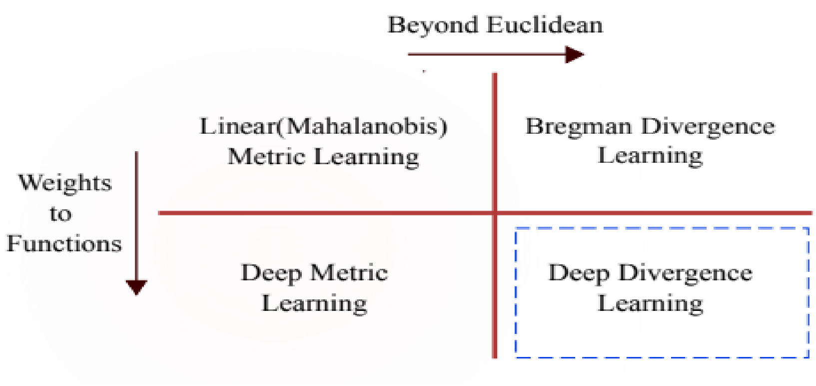


Figure 2.10: Classical linear metric learning methods (Cilingir et al. 2020)

2.6 Information Length Theory

Since Shannon entropy is not an ideal descriptor(Chamorro et al 2022) of the occurring statistical variations within a time series, this has created a great motivation for using other different information theoretic concepts , for example Fisher Information (Zegers 2015; Ly et al 2017; Frieden 2004), differential entropy (Michalowicz et al 2013), the Kullback-Leibler divergence (KD) (Van Erven and Harremos 2014), or the information length (IL) (Kim 2018; Kim and Hollerbach 2017). The total number of statistical variations for each specified temporal range can be calculated using IL. IL is superior to other information metrics like differential entropy since it depicts evolution path dependency between two states (PDFs) (Nicholson et al 2020). The beauty of IL metric is based on its capability to track the variability (Chamorro et al 2022) through the evolution of time series via time-dependent probability density functions (PDFs).

Additionally, IL's formalism introduces a fascinating connection between information geometry and stochastic processes (Heseltine and Kim 2019). More interestingly, IL has numerous applications for quantum, fluid, and biological processes (Kim 2019). On the other hand, IL metric was the real motivation behind the provision of a novel info-geometric measure of casual information rate (Kim and Guel-Cortez 2021). The significance of IL for stochastic thermodynamics is further discussed (Kim 2021), particularly in connection to its effects on the production of entropy or free energy in the non-autonomous Ornstein-Uhlenbeck process. To end, (Guel-Cortez and Kim 2020; Guel-Cortez and Kim 2021) present an analysis of the IL calculation of linear stochastic autonomous processes, which facilitates the range of applicability, facilitating application to various scenarios in engineering and demonstrating that IL can be employed for the abruption of event prediction.

2.6.1 IL as a concept

Mathematically speaking, if x serves as a nth- order stochastic variable and $p(x,t)$ is a time-dependent PDF of x, then the Information Length $\mathcal{L}(t)$ corresponding to its evolution from the initial time $t_0 = 0$ to the final time $t_F = t$ is devised by

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)} = \int_0^t \sqrt{(\varepsilon(t_1))}dt_1 \tag{2.36}$$

$$\varepsilon(t_1) = \int_{\mathbb{R}^n} \left(\frac{1}{p(x,t_1)}[\frac{\partial p(x,t_1)}{\partial t_1}]^2\right) dx \tag{2.37}$$

provided that $\sqrt{(}\varepsilon(t)$ serves as the root-mean- squared fluctuating energy rate.

Having a closer look at (2.36), it is essential to note that $\tau(t)$ serves as a dynamic temporal unit which provides the correlation time over which the changes of $p(x,t)$ take place (Nicholson et al 2020). Moreover, $\tau(t)$ acts as a time unit in the statistical space. Having said that, $\sqrt{(}\varepsilon(t) = \frac{1}{\tau(t)}$ is the quantification of the average rate of information’s change in time, or the information velocity(Kim 2021).

To understand the interpretation of $\mathcal{L}$, it is more favourable to calculate its underlying value of the corresponding physical process’ mathematical model. Considering the first order stochastic process described by Langevin equation:

$$\frac{dx}{dt} = -\gamma(t)\big(x - f(t)\big) + \xi \tag{2.38}$$

x serves as a random variable, f defines a deterministic force, ξ represents a short-correlated random force satisfying that:

$$< \xi(t)\xi(t_1) > = 2D\delta(t - t_1) \text{ and } < \xi(t) > = 0 \tag{2.39}$$

where D serves as the amplitude (temperature) of the deterministic force(stochastic nose) ξ

It is to be noted that equation (2.38) is so popular to be used as a descriptor of the motion of a particle under a harmonic potential in the form:

$$V(x) = \frac{1}{2}\gamma(x - f(t))^2 \tag{2.40}$$

Following (Heseltine and Kim 2016; Kim et al 2016), it is found that:

$$\varepsilon(t) = \frac{(\frac{d\beta}{dt})^2}{2\beta^2} + 2\beta(\frac{dy}{dt})^2 \tag{2.41}$$

$$p(x,t) = \sqrt{\frac{\beta}{\pi}} e^{-(\beta(x-y))^2}, y(t) = < x > = x(0)e^{-G(t)} + F(t)$$

$$\frac{1}{2\beta(t)} = < (x(t) - y(t))^2 > = \int_0^t 2D\, e^{-2(G(t)-G(t_1))} dt_1,$$

$$F(t) = \int_0^t e^{-\left(G(t)-G(t_1)\right)} \gamma(t_1) f(t_1) dt_1, \tag{2.42}$$

provided that $G(t) = \int_0^t 2D\, \gamma(t')dt'$.

Considering equation (2.38), it is clear that $\varepsilon(t)$ is dependent on the changes in both mean and variance defined by the corresponding dynamics of equation (2.33), portraying the changes of $p(x,t)$ in a three-dimensional space $(t, x, p(x,t))$ as in figure 2.11, where the variation of the information velocity, $\sqrt{(\varepsilon(t)}$ would occur along the path starting initially from the state probability density function $p(x, t_0)$ to the final state at $p(x, t_F)$ as a descriptor of the speed limit from the statistical deviations of the observables(Nicholson 2020). Therefore, the temporal integral of $\surd(\varepsilon(t)$ provides the changes in the mean y and the variance $\frac{1}{2\beta(t)}$ that take place along the path.

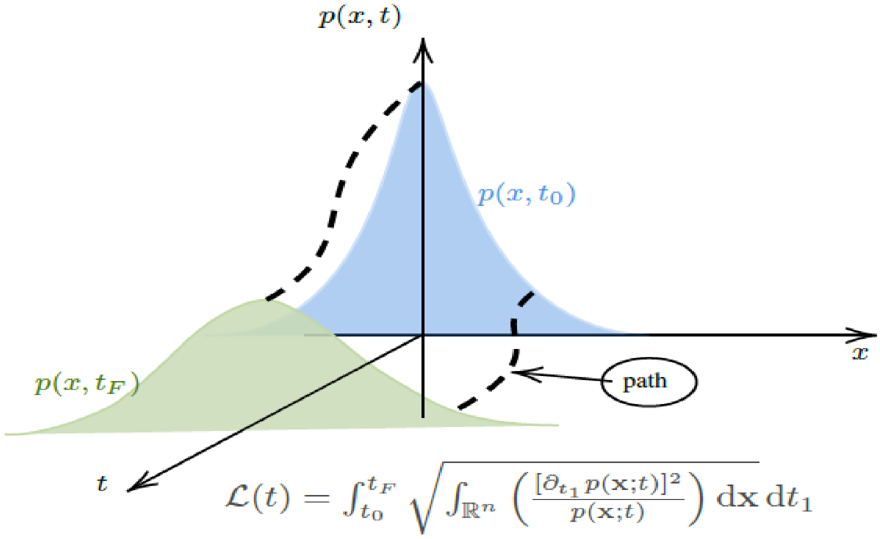


Figure 2.11: A graph depicting the evolution of $p(x, t)$ over time t. $\mathcal{L}(t)$ computes the total amount of statistical changes on $p(x, t)$ from t_0 to t_F (Chamorro and colleagues, 2022).

Fundamentally, employing differential entropy (Michalowicz 2013), may not enable us to observe the occurring temporal statistical variations. This is a direct implication of the locality's deficiency since differential entropy is mainly for the quantification of the differences between any two given PDFs disregarding any intermediate states (Heseltine and Kim 2019). In a different symbolism, it only notifies us of the differences that have an influence on the underlying system's general development. IL $\mathcal{L}(t)$, on the other hand, measures any localised changes that occur along the system's course (Kim 2018; Kim and Hollerbach 2017). More crucially, IL has been touted as a cutting-edge technique for depicting an attractor structure and as a potent metric that can unite geometry and stochasticity (Guel-Cortez and Kim 2020; Heseltine and Kim 2019). From the point of stable equilibrium, the equivalent value of $lim_{t\rightarrow\infty}\,\mathcal{L}(t)$ would grow linearly depending on where the starting state's mean PDF $p(x,0)$ is located (Kim and Hollerbach 2017; Hollerbach et al 2018).Notably, this strongly underlines that IL preserves the underlying Gaussian process' linear geometry. by utilising equation (2.36).More importantly, this particular property is lost when employing any other information metric(Guel-Cortez and Kim 2020 ;Heseltine and Kim 2019).By using definition, this emphasises that IL is a one-dimensional, model-free measure (2.36). Since IL is independent of data type, it may be devised by calculating the time-variant PDF of a time series (Chamorro et al 2022).

2.7 Existing research gaps

This thesis is aimed to fill several research gaps in the literature. To start with, the continuous time domain for the class of Rényi generalized entropies has left several unsolved open problems, such as the inability to explore many information-theoretic properties for this class. More fundamentally, this specific continuous time domain was impossible to extend the limit theorem beyond the origin point.

Notably, the Shannonian entropic formalism for the stable $M/G/1$ queue was the only known case in the literature since 1983. But what about the more generalized formalisms for Rényi and Tsallis entropic measures? This is a potential research gap in information-theoretic queueing theory.

Additionally, there is only a single paper in the literature on the info-geometric impact on the stable $M/D/1$ queue, which leaves the literature with numerous unanswered questions on the info-geometric analysis on stable queues.

Finally, the fact that there was no research in the literature on the information length theory of transient queues, has generated an exceptional motivation to undertake this novel research track.

To this end, the current thesis aims to answer the above-mentioned open problems to fill these research gaps.

2.8 Chapter summary

The current provides an overview of Rényi entropy and uncertain reasoning, followed by an introduction to the Maximum Entropy (ME) principle and its application to discrete ME distributions. It also summarizes ME solutions for queueing system performance distributions, stable queueing systems with two real -life applications of the transient $M/M/\infty$ queue. Additionally, the chapter reviews information geometry and information length theory, while highlighting existing research gaps and the aims and objectives of the study.

3. Properties of Discrete Rényi's Generalized Entropies Extended Properties and PV-updates

This chapter investigates the properties of Rényi's generalised entropies (RGEs) and their applications in information theory. It highlights that while these properties have been extensively studied in the continuous-time domain, this chapter presents an original extension of these properties into the discrete-time domain. It also provides Probability Vector Updates (PV-updates) and their connection to prior information-theoretic results on minimum cross entropy.

3.1 Introduction

One of the pillars of information theory is the investigation of the characteristics of Rényi's generalised entropies (RGEs) in both continuous and discrete time domains. They make it possible for more applications to be made in various domain in science and engineering. This includes inference and statistical mechanics, telecommunications networks, medicine, and other branches of human knowledge.

Several researchers have reported on the properties of RGEs on the continuous-time domain at the real line's origin (Kybic 2006). In addition, (Paris 1994) proposed inference processes based on the definition of RGEs, H_q^L in continuous-time domain, where H_q^L is Rényi's generalised entropy, L denotes a finite language (i.e., a finite set of propositional variables), and $q \in (-1, \infty)$ is the order of RGE (which is the parameter of the power in the definition of RGE). The theoretical result of RGEs, H_q^L, i.e., $\lim_{q \to a} H_q^L = H_a^L$, was only proved for order $a = 0$ (Uffink 1995). Hence, the usefulness of their associated proofs is of limited value.

More fundamentally, the current chapter deals with the extension of a general theorem for the class of RGEs in the discrete-time domain over real numbers, where the corresponding **Probability Vector Updates,** namely PV-updates of the class of RGEs are devised. Hence, a more general novel proof for all $a \in$ (–1,∞) is devised for the class of RGEs, H_q^L in the discrete time domain. See (Uffink 1995) for a more detailed account on these properties. Moreover, a generalisation of the results reported in (Paris and Vencovska 1992) is carried out, enabling the computation of the PV-updates with respect to the class of RGEs.

The current chapter focuses on providing comprehensive proofs for the extended properties of Rényi's Generalized Entropies, as well as deriving PV-updates in the Discrete Time Domain. Additionally, the chapter presents the full proofs of an expanded version of the limit theorem and offers physical interpretations for these extended properties. In principle, this chapter primarily contributes to employing the discrete case of RGEs to solve an open problem (Paris and Vencovska 1992) that has been reported to be unsolvable by using the continuous case, namely the limit theorem, ., $\lim_{q \to a} H_q^L \ = \ H_a^L$.

3.2 Background and definitions.

3.2.1 A closer look at uncertain reasoning

Rényi`s entropies encouraged the deduction of latest measures for assessing sign facts and intricacy within the time-recurrence plan. As implemented with a time–frequency representation (TFR) from the Cohen`s class (Boashash and Ouelha 2108; Jurdana et al. 2021) the Rényi's entropies alter closely to the concept of intricacy that it's far applied even as outwardly reviewing time-recurrence pictures. These actions have a few extra intriguing and helpful properties, like transformation invariances, cross-part and cross-component and accounting, that motivated researchers to choose them for time frequency analysis. A fundamental detailed investigation of the properties and a few expected uses of Rényi's entropies is accounted for in (Jurdana et al. 2021), with accentuation on the underlying mathematical framework for time frequency representations of the second order. Specifically, for Wigner's circulation, it was laid out that there exist signals for which the actions are not obvious (Boashash and Ouelha 2108). An extended entropy and mutual information estimator were announced in (Tarighi et al 2022) for Rényi's meanings of entropy, which includes those of Shannon's as exceptional cases.

Rather than utilizing the standard proportion of entropy (first proposed by Shannon), an action that exists in the group of Rényi's entropies was proposed in (Tarighi et al 2022) by permitting the likelihood evaluations to sometimes lie outside the reach from 0 to 1, to effectively expand Rényi's quadratic entropy corresponding to requirements communicated in the form of linear equations.

In a deeper insight, it is revealed by (Giovannetti et al 2004) that the minimization of Rényi's output entropy can be a consequence of coherent-state inputs. Looking at the more general case, namely arbitrary input states and non-integer orders. Moreover, it has been shown that the upper bound implied by the conjecture is compatible with the entropic lower bound obtained by (Giovannetti et al 2004). Furthermore, (Aggarwal 2005) suggested a different approach to using the Rényi's entropy to generalize the Tunstall codes (Tunstall 1968), which illustrates the influential role of Rényi's entropy by devolving an algorithmic minimization of redundancy. Additionally, the validity of Tunstall's theorem (Trunstall 1968) was extended to the countable alphabet by

(Aggarwal 2005) through examining the effects of two different definitions of mutual information on the generalisation of capacity and rate distortion function.

According to (Hawes 2007), the spectrum of Rényi's inference processes in discrete time has limits of minimax at one end and the limit centre of mass inference process CM_{∞}at the other. It was discovered that a different series of procedures had the limit *maximin*. Maximin is the dual of minimax. However, when compared to maximum entropy (ME), it exhibits traits that are superior to those of minimax. (Hawes 2007) provides algorithms for computing *minimax* and *maximin*, which have the benefit over those of ME in that they can infer belief values, which are rational numbers when the agent's knowledge is stated exclusively in terms of rational numbers.

There are several applications that use these estimators in the context of the nonparametric problem of estimating Rényi's entropy and mutual information (MI), based on a limited sample selected from an unknown, continuous distribution over $\mathbb{R}^d$ (the d- fold Cartesian product of real numbers $\mathbb{R}$). (Van Hulled 2008; Uffink 1995) have both employed entropy estimators and mutual information estimators for subspace analysis and picture registration, respectively. A class of estimators for the Shannon's and Rényi's data on multi-dimensional probability density was also offered by Leonenko (2010).

When developing an expert system, we need a rationale for picking an element from $V^L(K)$. Choosing a particular reasoning process N. For consistent (i.e., non-trivial) *K* [(i.e., $V^L(K) \neq \emptyset$], in practice it is very rare if $V^L(K)$is a singleton (i.e., a set with one element).

3.2.2. Definitions (Paris 1994)

1. For any finite language L, define a family of inference processes N^{L}. Within this context, $\mathrm{N}^{\mathrm{L}_2}$(K) agrees with $\mathrm{N}^{\mathrm{L}_1}$(K) on SL_1 for $\mathrm{L}_1 \subseteq \mathrm{L}_2$ (equivalently, $\mathrm{SL}_1 \subseteq \mathrm{SL}_2$ and $\mathrm{CL}_1 \subseteq \mathrm{CL}_2$). Additionally, K$\in \mathrm{CL}_1$.4. MD^{L} stands for minimum distance inference process, $\mathrm{MD}^{\mathrm{L}} = \text{that } \mathrm{x}^{\rightarrow} \in \mathrm{V}^{\mathrm{L}}(\mathrm{K})$ where $\sum_{\mathrm{i}=1}^{\mathrm{J}} \mathrm{x}_{\mathrm{i}}^2$ is minimal.

2. If maximum entropy is rewritten as an inference process , defined by ME^{L} , which is determined by

$\mathrm{ME}^{\mathrm{L}}(\mathrm{K}) = \text{that } x^{\rightarrow}$ satisfying the maximality of the entropy$-\sum_{i=1}^{J} x_i log x_i$ (Take

$xlogx = 0$ when $x = 0$).

3. If maximum entropy is rewritten as a transfer probability process, then we write it in short as MTP. This is read as:

$\mathrm{MTP}(\mathrm{K}) = \mathrm{x}^{\rightarrow} \in \mathrm{V}^{\mathrm{L}}(\mathrm{K})$ satisfying the maximality of $\sum_{\mathrm{i}=1}^{\mathrm{J}} \sqrt{\mathrm{x}_{\mathrm{i}}}$

4**.** Let G be the collection of constraints where $\{\mathrm{Bel}^{\sigma}(\alpha_{\mathrm{i}}) = 0\text{: i}\in I\}$, $I^L(K) = \{i: \forall x^{\rightarrow} \in \mathrm{V}^{\mathrm{L}}(\mathrm{K}),\ x_i = 0\}$. If $N^L =$ the maximal uniquely defined point $x^{\rightarrow} \in \mathrm{V}^{\mathrm{L}}(\mathrm{K})$ of F^{L},F^{L} serves as a function from $\mathbb{Q}^L$ into a set that is ordered linearly.This translates to the obstinacy of NL.

5**.**The minimum cross entropy update of the consistent K+K^{σ} with respect to an open-minded inference process N by choosing $< \mathrm{Bel}^{\sigma}(\alpha_1), \ldots, \mathrm{Bel}^{\sigma}(\alpha_J) >$ to be that $\mathrm{x}^{\rightarrow} \in \mathrm{V}^{\mathrm{L}}(\mathrm{K}^{\sigma})$ at which the function

$$I(\mathrm{x}^{\rightarrow}, N(K)) = \sum_{i=1}^{J} x_i \log(x_i / N(K)(\alpha_i)) = \sum_{i=1}^{J} (x_i \log(x_i) - x_i \log N(K)(\alpha_i)) \qquad (3.1)$$

is minimal. It is agreed that $0log0 = 0$.

3.3 Background theorems

3.3.1 Functional continuity theorem

For some continuous strictly concave(convex) function *F*: $\mathbb{Q}^L \rightarrow \mathbb{R}$, if the inference process *N* on *L* satisfies *N(K)* = that $x^{\rightarrow} \in V^L(K)$at which $F(x^{\rightarrow})$ is maximal(minimal), then *N* is continuous (PARIS 1994).

3.3.2. Properties of Maximum Entropy ME

It has been conjectured (Paris 1994) that only ME is continuous, obstinate, open-minded. Additionally, ME satisfies the relativization, independence, renaming and irrelevant information (Paris 1994). (Paris 2003) used a novel approach by incorporating a "common sense principle" to provide a second reasoning for this choice. As an illustration, let's say a scientist was shown a slide under the microscope, that was unquestionably displaying a follicular pattern, this implies Bel^{σ}(F) =1, where $\mathrm{Bel}^{\sigma}(\theta)$ for θ ∈SL represents the scientist's conviction that σ has a property θ.

In this manner, the new, unique, beliefs about σ obtain similar status as the overall information explanations. In broad terms, an expert now defines constraints K^{σ} on Bel^{σ}. Consequently, we must upload to K constraints, every $\mathrm{Bel}^{\sigma}(\theta)$ change through Bel $((s\wedge\theta)/\mathrm{Bel}(s)$. Thus, it appears logical to keep in mind the updated belief that σ has a property θ, known as $\mathrm{Bel}^{\sigma}(\theta)$, to be:

$$\mathrm{N}(\mathrm{K}+\mathrm{K}^{\sigma}(S))(S\wedge\theta)/\,\mathrm{N}(\mathrm{K}+\mathrm{K}^{\sigma}(S))(S) \qquad (3.2)$$

Tragically, K+$\mathrm{K}^{\sigma}(S)$ presently excludes one significant extra conviction σ, that ‘being like σ 'is impossible. To be sure, belief in it ought to be basically 'infinitesimal' given that one is sufficiently hard. This exclusion can create perplexing outcomes in small models, albeit in enormous, true models it would make an irrelevant difference. To address this, we add

$$\mathrm{Bel}(S) = \varepsilon \text{ to } \mathrm{K}+\mathrm{K}^{\sigma}(S) \qquad (3.3)$$

3.4 Extended Properties for RGEs into the Discrete-time Domain

This current section extends the proofs of several properties of RGEs from the continuous case into the discrete case. In this context, the main problem of the study is composed of the following mechanism: For $q > -1, q \neq 0$, we write $H_q^L(K)$ as

$$H_q^L(K) = \text{ that } \mathrm{x}^{\rightarrow} \in V^L(K) \text{ for which } (\textstyle\sum_{i=1}^{J} x_i^{q+1})^{-1/q} \text{ is maximal} \tag{3.4}$$

Also, we define U_0^L to be ME. We start by giving the definition of each property given, along with its physical interpretation and investigation of these properties of the RGE in the discrete-time domain. Salient proofs are introduced, and the full proofs are detailed in the Appendices chapter.

3.4.1 Uniqueness property

Essentially, on the premise of this saying, "If a similar issue is tackled two times similarly, we expect similar response in the two cases i.e., the arrangement ought to be clear" (Rödder 2019). Thus, we can compose this as:

H_q^L is unique if there exits $\in V^L(K)$ satisfying the maximality of $(\sum_{i=1}^{J} x_i^{q+1})^{-1/q}$

The salient proof of 3.4.1

The proof covers three possibilities:

First possibility: $q > 0$.

Second possibility: $q = 0$.

Third possibility: $0 > q > -1$

The sketch of proof of the first case starts by showing the convexity of x^{q+1} on [0,1]. The second step is to show the convexity of $\sum_{i=1}^{J=2^n} x_i^{q+1}$ on $\mathbb{Q}^L$, meaning that for all 0 $\le \lambda \le$1, i= 1, ..., J = 2^n, $a^{\rightarrow}, b^{\rightarrow} \in \mathbb{Q}^L$,

$$\sum_i(\lambda a_i + (1-\lambda)b_i)^{q+1} \le \lambda \sum_i a_i^{q+1} + (1-\lambda) \sum_i b_i^{q+1} \tag{3.5}$$

Which proves the required result by contradiction. By H_0^L = ME, case 2 is immediately proved (Paris 2003). The proof of case 3 is like that of case 1. The full proof of uniqueness Property can be seen in the Appendices chapter.

3.4.2 The Null Limit Theorem

$$\lim_{q\to 0} H_q^L = \text{ME} \tag{3.6}$$

The proof starts by showing the : (i) Uniform convergence i.e., that if given $\varepsilon > 0$ $\exists \delta > 0$ such that if

$$|q| < \delta, x^{\rightarrow} \in \mathbb{Q}^L, \text{ then } |(1/q)log(\sum_{i=1}^{J} x_i^{q+1}) - \sum_{i=1}^{J} x_i log x_i| < \varepsilon \tag{3.7}$$

(ii) If $a^{\rightarrow(q)} = < a_1^{(q)}, \ldots, a_J^{(q)} >$ is the point in $V^L(K)$ at which $log\left(\sum_{i=1}^{J} x_i^{q+1}\right)^{-\frac{1}{q}}$ is its maximum, then $lim_{q\to 0} a^{\rightarrow(q)} = a^{\rightarrow(0)}$.

Part (i) was shown by using the Taylor expansion for $og\left(\sum_{i=1}^{J} x_i^{q+1}\right)^{-\frac{1}{q}}$ around $q = 0$. Then the proof continues to show that: $[d^2/dt^2 log(\sum_{i=1}^{J} x_i^{t+1})]_{t=\theta_{x^{\rightarrow}}}$ has an independent of $x^{\rightarrow}$ upper bound, which was carried out with the help of elementary mathematical analysis. In a similar fashion, the proof can be devised for $q < 0$.

As for (ii), let's propose the contradicting statement , i.e., ∃a subsequence $a^{\rightarrow(q_n)}$ when $q_n \searrow 0$ such that $\lim_{n\to\infty} a^{\rightarrow(q_n)} = b^{\rightarrow} \neq a^{\rightarrow(0)}$. Since $b^{\rightarrow} \neq a^{\rightarrow(0)}$, ∃a positive number $\eta \in (0,1)$ such that $\sum_i b_i \, log b_i - \sum_i a_i^{(0)} \, log a_i^{(0)} > \eta > 0$ and $b^{\rightarrow} \in \mathrm{V^L(K)}$(Since

$V^L(K)$ is compact). Following this argument leads to $1\,/|q_n lo\,g\left(\sum_i (a_i^{(0)})^{1+r_n}\right)| <$ $1\,/|q_n \log\left(\sum_i (a_i^{(0)})^{1+q_n}\right)|$. This provides the converse argument. Following the same analogy, we can obtain the proof for $q_n \nearrow 0$.

Proof: See Appendices chapter for full proof.

3.4.3 The Limit Theory

$$\lim_{q \to a} H_q^L = H_a^L \qquad \forall -1 < a < \infty \tag{3.8}$$

The proof is obtained in the following sequence. Firstly, we have shown the uniform convergence. In other words, if given $\varepsilon > 0\ \exists \delta > 0$ such that if $|q - a| < \delta,$ then

$$|\sum_{i=1}^{J} x_i^{q+1} - \sum_{i=1}^{J} x_i^{a+1}| < \varepsilon \quad \text{for all } x^{\rightarrow} \in \mathbb{Q}^L \tag{3.9}$$

This was done by using the Taylor expansion for $\sum_{i=1}^{J} x_i^{q+1}$ around $q = a$ and showing that $\sum_{i=1}^{J} x_i^{a+1}\ log x_i$ and $[\sum_{i=1}^{J} x_i^{t+1}\ (log x_i)^2\ ]t = \theta_{x^{\rightarrow}}$ has an independent of $x^{\rightarrow}$ upper bound. A similar argument works for $q < a$. Secondly, we proved that if $b^{\rightarrow(q)} = < b_1^{\rightarrow(q)}, \ldots, b_J^{\rightarrow(q)} >$ is the point in $V^L(K)$ at which $\sum_{i=1}^{J} x_i^{q+1}$ is maximal (minimal), then $lim_{r \to a}\ b^{\rightarrow(q)} = b^{\rightarrow(a)}$. This is proven by contradiction by assuming the existence of a subsequence $b^{\rightarrow(q_n)}$ when $q_n \searrow a$ such that $lim_{n \to \infty}\ b^{\rightarrow(q_n)} = c^{\rightarrow} \neq b^{\rightarrow(a)}$, $\exists$ a positive number $\eta \epsilon\ (0{,}1)$ such that we have $\sum_i {c_i}^{a+1} - \sum_i (b_i^{(a)})^{a+1} > \eta > 0$ and $c^{\rightarrow} \in V^L(K)$ (Since $V^L(K)$ is compact). Putting all the threads together leads to the required contradiction:

$$\sum_i (b_i^{(a)})^{1+q_n} < \sum_i (b_i^{(q_n)})^{1+q_n} \tag{3.10}$$

A similar argument works for $q_n \nearrow a$.

For the full proof please see the Appendices chapter.

3.4.4. Extended properties of RGEs in the discrete case

3.4.4.1. The physical interpretation of irrelevant information

Suppose that $K_1, K_2 \in \mathrm{CL}$, $\theta \in \mathrm{SL}$, but either no propositional variable appears in θ or any sentence in K_1 additionally appears in K_2 .Then $N(K_1 + K_2)(\theta) = \mathrm{N}(K_1)(\theta)$ This principle may well be physically translated as knowledge entirely irrelevant to the problem in hand can be ignored (Hawes 2007).

Firstly, H_0 = ME satisfies irrelevant information by equation (3.6). Let $q > 0$ and K_1be the set of constraints $x + y + z = 1(so\ Bel(\neg p_1 \wedge \neg p_2) = 0\ automatically)$ $Ny + z = f$ where $N > 1$ is large and $f = \frac{1}{1+(\frac{N}{N-1})^{1/q}}$ .From which it follows that

$$U_q^L(K_1)(p_1 \wedge \neg p_2) = 0 \qquad (3.11)$$

Letting

$$x_1 = Bel(p_1 \wedge p_2 \wedge p_3), y_1 = Bel(p_1 \wedge \neg p_2 \wedge p_3), z_1 = Bel(\neg p_1 \wedge \neg p_2 \wedge p_3), x_2 = Bel(p_1 \wedge p_2 \wedge \neg p_3), y_2 = Bel(p_1 \wedge \neg p_2 \wedge \neg p_3), z_1 = Bel(\neg p_1 \wedge \neg p_2 \wedge \neg p_3), \qquad (3.12)$$

and K_2be the set of constraints K_1+ $Bel(p_3) = d$. As we go along the proof, by assuming that U_q^L satisfied irrelevant information. Then, by language invariance,

$$H_q^L(K_2)(p_1 \wedge \neg p_2) = H_q^L(\mathrm{p}(p_1 \wedge \neg p_2) = 0 \qquad (3.13)$$

Therefore, by obstinacy $H_q(K_2) = H_q(K_3)$ where K_3 is the set of constraints:

$y_1 + y_2 = 0\ giving\ z_1 = f - z_2$, $x_1 + x_2 + z_1 + z_2 = 1\ giving\ x_1 = d - f + z_2$,$z_1 + z_2 = f$, $x_1 + z_1 = d\ giving\ x_2 = 1 - d - z_2$. From which it follows few steps of the proof that if $\varepsilon > 0$ is very small and

$$x_1' = d - f + (N-1)\varepsilon, x_2' = 1 - d, y_1' = \varepsilon, y_2' = 0,\ z_1' = f - N\varepsilon, z_2' = 0 \qquad (3.14)$$

We get a contradiction. In other words, H_q^L cannot satisfy irrelevant information. Finally, we consolidate our proof by setting a counter example to show that $H_{-1/2}$ fails to satisfy irrelevant information.

Check the Appendices chapter for the more detailed proof.

3.4.4.2. The open –mindedness property

Define K $\in \mathrm{CL}, \theta \in$ SL such that K+Bel$(\theta) \neq 0$ is consistent, then $N(K)(\theta) \neq 0$

This is justified (Hawes 2007) by assuming that our knowledge does not compel us to be in a situation when Bel (θ) = 0, then we may not infer that belief as this would falsify θ unnecessarily. Therefore, if accepting that Bel (θ) = 0 is an extreme view, then open-mindedness is a favourable property of an inference process.

Proof:

We know that $H_0^L = ME$ is open-minded by (3.3.2). For $-1 < q < 0$, suppose

$$H_q^L(K) = \rho^{\rightarrow},\ H_q^L(K)(\theta) = 0 \tag{3.15}$$

whilst there is $a^{\rightarrow} \in V^L(K)$ with $a_j > 0$ for $\alpha_j \in S_\theta$. Then for ε small,

$$\rho^{\rightarrow} + \varepsilon(a^{\rightarrow} - \rho^{\rightarrow}) \in V^L(K) \tag{3.16}$$

and by the choice of $\rho^{\rightarrow}$,

$$\sum_i \rho_i^{q+1} - \sum_i(\rho_i + \varepsilon(a_i - \rho_i))^{q+1} > 0 \tag{3.17}$$

Implies

$$(q+1)\varepsilon \sum_i(a_i - \rho_i)\big(\rho_i + \lambda\varepsilon(a_i - \rho_i)\big)^q < 0 \tag{3.18}$$

for some $0 < \lambda < 1$, by the mean value theorem. (3.18) could be written in the form:

$$0 > \left[\sum_{\alpha_i>0,\rho_i=0}(a_i-\rho_i)\big(\rho_i+\lambda\varepsilon(a_i-\rho_i)\big)^q + \sum_{\rho_i>0}(a_i \;-\rho_i)\big(\rho_i+\lambda\varepsilon(a_i-\rho_i)\big)^q\right] \tag{3.19}$$

The second term in (3.19) is bounded as $\varepsilon \searrow 0$. To see this we must notice that:

$$|\sum_{\rho_i>0}(a_i-\rho_i)\big(\rho_i+\lambda\varepsilon(a_i-\rho_i)\big)^q| \leq \sum_{\rho_i>0}|(a_i-\rho_i)|(\rho_i+\varepsilon)^q$$

$$\leq \sum_{\rho_i>0}|(a_i-\rho_i)|(\rho_i\,/\,2)^q$$

$$\leq \sum_{\rho_i>0}(3\rho_i/2)^q \;\; (for\ sufficiently\ small\ \varepsilon\,) \tag{3.20}$$

(Since $|(a_i-\rho_i)| \leq 1$ holds for all i =*1*, 2,…, *J*). On the other hand, the first sum in (3.19) equals $\sum_{\alpha_i>0} a_i(\lambda\varepsilon a_i)^q$which tends to infinity as ε tends towards zero (since there is at least one such a_i). Therefore, the right-hand side of (3.19) tends to infinity whenever $\varepsilon \searrow 0$), giving the required contradiction.

<u>Counter example to open-mindedness for *q* > 0</u>

Put $x_i = Bel(\alpha_i)$for *i* =1,2,3,.., 2^n and

$$K = \begin{cases} x_i = 0 & i = 4,5,..,2^n \\ x_2 = \varepsilon - 4\varepsilon x_1 \end{cases} \tag{3.21}$$

where $\varepsilon > 0$ is very small (just how small will depend on *r*). The possible range of values of x_1 is [0,1/ 4]. Now since

$$x_1 + x_2 + x_3 = 1, x_3 = (1-\varepsilon) + x_1(4\varepsilon - 1) \tag{3.22}$$

So on K,

$$x_1^{q+1}+x_2^{q+1}+x_3^{q+1} = [x_1^{q+1} + (\varepsilon - 4\varepsilon x_1\,)^{q+1} + x_3^{q+1}\;] = g(x_1) \tag{3.23}$$

Differentiating g with respect to x_1 gives

$$dg/dx_1 = (q+1)[\,x_1^q - 4\varepsilon(\varepsilon - 4\varepsilon x_1\,)^q - (1-4\varepsilon)\,((1-\varepsilon) + x_1(4\varepsilon-1))^q] < 0 \tag{3.24}$$

since for $x_1 \in [0,1/4]$

$$x_1^q \leq (1\,/\,4)^q < (1-4\varepsilon)\,((1-\varepsilon) - 1\,/\,4(1-4\varepsilon))^q \;\; (\text{recall } \varepsilon\ is\ very\ small\,)$$

$$\leq \ (1\,/\,4)^q < (1-4\varepsilon)\,((1-\varepsilon) - x_1(1-4\varepsilon))^q \quad (3.25)$$

It follows that the value of $H_q^L(K)(\alpha_1)$ will be $1\,/\,4$, and hence $H_q^L(K)(x_2) = 0$, contradicting open-mindedness.

3.4.4.3. The physical interpretation of renaming principle

Suppose $K_1, K_2 \in \mathrm{CL}$,

$$K_1 = \{\sum_{j=1}^{q} a_{ji}\, Bel\,(\gamma_j) = b_i | i = 1,2,\ldots,m\} \quad (3.26)$$

where a_{ji} , b_i are real and $\theta_j \in \mathbb{R}$.

$$K_2 = \{\sum_{j=1}^{q} a_{ji}\, Bel\,(\delta_j) = b_i | i = 1,2,\ldots,m\} \quad (3.27)$$

where $\gamma_1,\ldots,\gamma_J$, $\delta_1,\ldots,\delta_J$ are permutations of $\alpha_1,\ldots,\alpha_J$. Then,

$$N(K_1)(\gamma_j) = \mathrm{N}(K_1)(\delta_j) \quad (3.28)$$

Simply means that changing the names we call things should not change the probabilities we assign to them (Hawes 2007).

Proof:

Suppose $K_1, K_2 \in CL$,

$$K_1 = \{\sum_{j=1}^{2^n} a_{ji}\, Bel(\alpha_j) = b_i |\ i = 1,\ldots,m\}, K_2 = \{\sum_{j=1}^{q} a_{ji}\, Bel(\gamma_j) = b_i |\ i = 1,\ldots,m\} \quad (3.29)$$

where γ_j are permutations of the atoms α_j of L. We prove that:

$$H_q^L(K_1)(\alpha_j) = H_q^L(K_2)(\gamma_j),\ j = 1,2,3\ldots,\ 2^n \quad (3.30)$$

We have

$$H_q^L(K_1) = that\ < x_1, x_2, \ldots, x_{2^n} > \quad (3.31)$$

such that $x_j \geq 0, \sum_{i=1}^{2^n} a_{ji}\, x_j = b_i$ for $i = 1,\ldots,m$ for which $(\sum_{j=1}^{J}(x_j^{q+1})^{\frac{-1}{q}})$ is maximal.

We know that γ_j is a permutation of α_j, in other words, $\gamma_j = \alpha_{\sigma(j)}$, for some permutation σ of $\{1,2,3,..,2^n\}$. Hence

$H_q^L(K_1)$ = that $< x_1, x_2, \ldots, x_{2^n} >$ such that $x_j \geq 0$, $\sum_{i=1}^{2^n} a_{ji}\, x_{\sigma(j)} = b_i$ for $i = 1, m$ for which $(\sum_{j=1}^{J}(x_{\sigma(j)}^{q+1})^{\frac{-1}{q}})$ is maximal, which is essentially the same thing.

Therefore, the vector $< x_1, x_2, \ldots, x_{2^n} >$ such that

$$H_q^L(K_2) = < x_{\sigma^{-1}}(1), x_{\sigma^{-1}}(2), \ldots, x_{\sigma^{-1}}(2^n) > = H_q^L(K_1) \tag{3.32}$$

From which it follows that $H_q^L(K_1)(\alpha_j) = H_q^L(K_2)(\alpha_{\sigma(j)})$. QED

3.4.4.4. The obstinacy principle

Suppose $K_1, K_2 \in \mathrm{CL}$ and $N(K_1)$ satisfies K_2. Then:

$$N(K_1 + K_2) = \mathrm{N}(K_1) \tag{3.33}$$

Getting evidence to back up what we already think shouldn't change our minds (Hawes 2007).

Clearly obstinacy of H_q^L holds as it is a maximum inference process(c.f., part 7 of definition (3.2.2).

3.4.4.5. The relativisation principle

Suppose $K_1, K_2 \in CL$, $0 < c < 1$ and

$$K_1 = \{Bel(\emptyset) = c\} + \{\textstyle\sum_{j=1}^{q} a_{ji}\, Bel\,(\theta_i/\emptyset) = b_i : i = 1,2,\ldots,m\} \tag{3.34}$$

and

$$K_2 = K_1 + \{\textstyle\sum_{j=1}^{q} a_{ji}\, Bel\,(\delta_j) = b_i : i = 1,2,\ldots,m\} \tag{3.35}$$

Then for $\theta \in SL$,

$$N(K_1)(\theta/\emptyset) = N(K_2)(\theta/\emptyset) \tag{3.36}$$

The probability one would assign to such events should only be based on the knowledge one would have in certain circumstances (Hawes 2007).

The salient proof of 3.4.4.5

The proof starts by perfectly choosing the sets K_1,$K_2 \in CL, 0 < c < 1$, $K'_2 = K'_1$, K''_1, and K''_2. The proof continues to show the consistency of K''_2. Having this done, the proof goes further to investigate the following cases:

Case 1: $\infty > q > 0$.The proof of case 1 is carried out by showing:

$$E(x^{\rightarrow}) \geq E(c\rho^{\rightarrow} + (1-c)\tau^{\rightarrow}) \geq E(\nu^{\rightarrow}) \tag{3.37}$$

which implies, $\nu_i = \rho_i c$ for $i = 1, \ldots, h$. Following a similar argument shows that:

$$H_q^L (K_1)(\theta \wedge \phi) = H_q^L (K_2)(\theta \wedge \emptyset) \quad \text{for } \theta \in SL \tag{3.38}$$

and the result follows. The remaining cases for $q = 0$ and $-1 < q < 0$ are immediate. For the detailed lengthy proof of (3.4.4.5), the reader is kindly advised to see the full proof in Appendices chapter.

3.4.4.6. Principle of independence

In the special case of $\{p_1, p_2, p_3\}$ and $K = \{Bel(p_1) = a, Bel(p_2/p_1) = b, Bel(p_3/p_1) = c\}$, whenever $a > 0$, if

$$N^L(K)(p_2 \wedge p_3/p_1) = b\,c = N^L(K)(p_2/p_1).N^L(K)(p_3/p_1) \tag{3.39}$$

then $N^L(K)$ satisfies the principle of independence.

In mathematical logic, independence refers to the improvability of a sentence from other sentences (Hawes 2007).

The salient proof of 3.4.4.6

The proof starts by considering the set:

$$K = \{(Bel(p_1) = 1, Bel(p_2 / p_1) = b, Bel(p_3 / p_1) = b\} \tag{3.40}$$

and by showing that for, then:

$$H_q^L(K)(p_2 \wedge p_3|p_1) \neq b^2 \tag{3.41}$$

will hold for the following cases:

1. $q = 0$.   2. $-1 < q,\ q \neq 0$

The first case is immediate by theorem (3.4.2)

As for the remaining cases, assume $q > 0$ (the case for $-1 < q < 0$ is similar) and let:

$$x_1 = Bel(p_1 \wedge p_2 \wedge p_3) \tag{3.42}$$

$$x_2 = Bel(p_1 \wedge p_2 \wedge \neg p_3) \tag{3.43}$$

$$x_3 = Bel(p_1 \wedge \neg p_2 \wedge p_3) \tag{3.44}$$

$$x_4 = Bel(p_1 \wedge \neg p_2 \wedge \neg p_3) \tag{3.45}$$

,x_5, x_6, x_7, x_8 for the remaining atoms, would yield after lengthy calculations,

$V^L(K) = \{< x_1, b - x_1, b - x_1, 1 - 2b + x_1, 0,0,0,0 > | 1 \geq x_1 \ \geq 0, 1 \geq 1 - 2b + x_1 \ \geq$

$0,1 \geq b - x_1 \ \ \geq 0\}$ (3.46)

Assuming $b < \frac{1}{2}$ and sorting out the inequalities defining V^L(K) implies

$$V^L(K) = \{< x_1, b - x_1, b - x_1, 1 - 2b + x_1, 0,0,0,0 > |\ x_1 \ \in [0, b]\} \tag{3.47}$$

The proof proceeds to compute $\mathrm{x}^{\rightarrow} \in V^L(K)$ at which $(x_1) = \sum_{i=1}^{J} x_i^{q+1}$is minimum. Carrying out the steps of calculations shows that $x_1 = b^2$ will give a minimum value of $f_q < x_1, x_2, \ldots, x_J >$. This implies:

$(b^q)^2 - 2b^q(1-b)^q + ((1-b)^q)^2 \Leftrightarrow (b^q - (1-b)^q)^2 = 0$

$$\Leftrightarrow (b^q - (1-b)^q = 0 \Leftrightarrow b = \frac{1}{2} \tag{3.48}$$

which is a contradiction. So, since $0 < b^2 < b$ (recall the range of x_1 here is [0,b]) this cannot not be a minimum point of $f_q(x_1^{\rightarrow})$. Therefore, we have

N^L(K)$(p_2 \wedge p_3 | p_1) \neq H_q^L(K)(p_2 | p_1)$. H_q^L(K)$(p_3 | p_1)$ = b^2 (Since H_q^L(K)(p_1) = 1) (3.49)

Hence, H_q^L does not satisfy independence at q = 0. The full proof is found in the Appendices chapter.

3.4.4.7. Principle of language invariance

Let L be a finite language and N^L be a family of inference processes. Then, N^L is language invariant if $L_1 \sqsubseteq L_2$ (so $SL_1 \sqsubseteq SL_2$ and $CL_1 \sqsubseteq CL_2$)and $K \in C\ L_1$ implies that $N^{L_2}(K)$ agrees with $N^{L_1}(K)$ on SL_1. The motivation behind language invariance is that this principle may be of interest to the finite language L at any time, but behind the fact that a rational choice of beliefs in that language should be able to extend to larger languages. Finally, there is no clear reason to assume that the relationships are finite, and L already contains them all. (Hawes 2007).

Proof

It is done if we show:

$$H_q^L(K)(\alpha_i) = H_q^{L'}(K)(\alpha_i) \tag{3.50}$$

where $L = \{p_1, \ldots, p_n\}$, $L' = L \cup \{p_{n+1}\}$, $K \in CL$ and, as usual, $At^L = \{\alpha_1, \ldots, \alpha_J\}$. Let $\beta_{2j-1} = \alpha_j \wedge p_{n+1}$ and $\beta_{2j} = \alpha_j \wedge \neg p_{n+1}$, so the β_{2j} for k =1,...,$2J$ are the atoms of L'. Now we notice that by the way $V^L(K)$ is formed,

$$< x_1, \ldots, x_{2J} > \in V^{L'}(K) \Leftrightarrow < x_1 + x_2, \ldots, x_{2J-1} + x_{2J} > \in V^L(K) \tag{3.51}$$

So $I^{L'}$(K) = {2j–1,2j|$j \in I^L(K)$}. Now, let us suppose that $H_q^L(K) = < \tau_1, \tau_2, \ldots, \tau_J > \in V^L(K)$,

$H_q^{L'}(K) = < \rho_1, \rho_2, \ldots, \rho_{2J} > \in V^{L'}(K)$. Then, by the definition, we need to study the following cases:

Case 1: $q > 0$. Case 2: $q = 0$. Case 3: $0 > q > -1$

As for case 1, we have by the definition:

$$\left(\sum_j \tau_j^{q+1}\right)^{-1/q} \geq \left(\sum_j (\rho_{2j-1} + \rho_{2j})^{q+1}\right)^{-1/q} (\text{By} < \rho_1 + \rho_2, \ldots, \rho_{2J-1} + \rho_{2J} > \in V^L \tag{3.52}$$

or equivalently

$\sum_j \tau_j^{q+1} \leq \sum_j \left(\rho_{2j-1} + \rho_{2j}\right)^{q+1} = 2^{q+1}\left(\sum_j \left(\rho_{2j-1}/2 + \rho_{2j}/2\right)^{q+1}\right.$

$\leq\ 2^{q+1}(1/2\,.\sum_j\left(\rho_{2j-1}\right)^{q+1} + 1\,/\,2.\sum_j(\rho_{2j})^{q+1})$ (since x^{q+1} is convex whenever $q\ > 0$)

$\leq 2^{q+1}(1/2\,.\sum_j\left(\tau_j\,/\,2\right)^{q+1} + 1\,/\,2.\sum_j(\tau_j\,/\,2)^{q+1})$(since $H_q^L(K) =\ x^{\rightarrow} \in$

$V^L(K)$ for which $\sum_{i=1}^{J} x_i^{q+1}$ is minimum, where $q > 0, < \frac{\tau_1}{2}, \frac{\tau_1}{2}, \frac{\tau_2}{2}, \frac{\tau_2}{2}, \ldots, \frac{\tau_J}{2}, \frac{\tau_J}{2} >$ $\in V^{L'}(\mathrm{K}))$

$$\leq 2^{q+1}(2.1\,/\,2.\sum_j\left(\tau_j/2\right)^{q+1}) = \sum_j {\tau_j}^{q+1} \tag{3.53}$$

Hence by the uniqueness of the minimum points:

$$(\rho_{2j-1} + \rho_{2j})\,/\,2 = \rho_{2j-1} = \rho_{2j} = \tau_j/2 \tag{3.54}$$

and the result follows since for $q > 0$, one gets:

$$H_q^L(\mathrm{K})(\alpha_i) =\ \tau_i = \rho_{2i-1} + \rho_{2i} \quad = H_q^{L'}(\mathrm{K})(\beta_{2i-1}) + H_q^{L'}(\mathrm{K})(\beta_{2i}) \quad = H_q^{L'}(\mathrm{K})(\alpha_i) \tag{3.55}$$

As for case 2, we know that H_0^L(K) = ME which is language invariant by theorem (3.4.2). Finally, the proof of case 3 is like case 1. So, we have H_q^L a language invariant for all $q\ > -1$.

3.4.4.8. The principle of continuity

The Blaschke metric(Pearl 1990), Δ is defined by Δ(C,D) = inf$\{\delta{:}\forall x^{\rightarrow} \in \mathrm{C}\ \exists y^{\rightarrow} \in \mathrm{D}, \mathrm{I}x^{\rightarrow} - y^{\rightarrow}\mathrm{I} \leq \delta\ \&\ \forall y^{\rightarrow} \in \mathrm{C}\ \exists x^{\rightarrow} \in \mathrm{D}, \mathrm{I}x^{\rightarrow} - y^{\rightarrow}\mathrm{I} \leq \delta\}$, with I$x^{\rightarrow} - y^{\rightarrow}$I to define the usual Euclidean distance between the two points $x^{\rightarrow}, y^{\rightarrow}$ and any arbitrary convex subsets C, D of the set $\mathbb{Q}^L$ = { $x^{\rightarrow} \in \mathbb{R}^J{:}\ x^{\rightarrow} \geq 0, \sum_{i=1}^{j} x_i$ =1}.The continuity requirement (Pearl 1990) that an inference process *N*, as a function of the convex set $V^L(K)$, is continuous with respect to this Blaschke metric, i.e., if $\theta \in SL, K, K_i\ \in CL$ for $i \in\ \mathbb{N}$ and $\lim_{i\to\infty} \Delta\left(V^L(K), V^L(K_i)\right) = 0$ then $\lim_{i\to\infty} N(K_i)(\theta) = N(K)(\theta)$. It makes sense that an inference process must satisfy the continuity concept, meaning that a microscopic change in the inferences is independent of a microscopic change in the information. One could contend that the rational agent's knowledge is positively varying a bit and that it would be absurd for those variations to result in appreciable changes in the belief inferred. Finding appropriate topology foundations of knowledge addressing the query "when are knowledge bases adjacent to each other?" is challenging when specifying this attribute. Saying that any two knowledge bases, K_1 and K_2, are closed if and only if their coefficients in the constraints are near is an obvious initial effort. Let them each have a matrix of coefficients A and B, and let's say that c_{ji}. Assuming that D is the matrix of knowledge base that was near C, it should not be close to A. Therefore, knowledge content must be considered in our concept of closeness (Hawes 2007).

Proof:

In case 3, within the proof of the uniqueness property (c.f.,3.4.1), we have $H_q^L(K) =$ that $\mathrm{x}^{\rightarrow} \in V^L(K)$ for which the maximality of $(\sum_i x_i^{q+1})$ combined with the strictly concaveness of the function $(\sum_i x_i^{q+1})$ hold. So, by (3.4.1), we have H_q^L is continuous for $0 > q > -1$. From the definition, we have H_0^L = ME which is continuous by the functional continuity theorem (c.f., 3.3.1). Finally, we have already proved that for $q > 0$, then $H_q^L(K) =$ that $\mathrm{x}^{\rightarrow} \in V^L(\mathrm{K})$ for which the maximality of $(\sum_i x_i^{q+1})$ as well as strictly concaveness of the function $(\sum_i x_i^{q+1})$ are satisfied. So, by (3.4.2), we conclude that $H_q^L(\mathrm{K})$ is continuous for $q > 0$. Therefore, $H_q^L(\mathrm{K})$ is continuous for all $q > -1$. Now, we come to introduce another contribution in the current paper, which is finding the PV-updates of the family of RGEs in the case.

3.5. Finding PV-Updates

Belief functions provide an approach that assimilates the influence of new evidence into a state of partial knowledge or partial belief. Encoding the initial state as a belief function and the evidence as a belief function, then the updated state of belief, accounting for the impact of the new evidence (Pearl 1990).

Define the PV-update of K+$K^{\sigma}(S)$ with respect to an inference process N by

$$Bel^{\sigma}(\theta) = \lim_{\varepsilon \searrow 0} N(K+K^{\sigma}(S)) + (Bel(S) = \varepsilon)(\theta|S) \tag{3.56}$$

For a more detailed survey on PV- updates, the reader is advised to see (Paris 1994; Paris &Vencovska1992). Recalling the definition of PV- Updates, we must clarify why we study PV-updates? The importance of PV-updates appears clearly from the findings approached by (Paris 1994), which are as follows:

The consistency for ε small follows from the consistency of K+ $K^{\sigma}(S)$ and that for *N* = *ME* the limit

$$\mathrm{Bel}^{\sigma}(\theta) = \lim_{\varepsilon \searrow 0} \mathrm{N}(\mathrm{K}+\mathrm{K}^{\sigma}(S)) + (\mathrm{Bel}(S) = \varepsilon)(\theta|S) \tag{3.57}$$

Always exists and, furthermore, in this case PV-updating agrees with minimum cross entropy updating (with respect to ME) and if we accept the earlier arguments for the choice of the maximum entropy inference process and we further accept the reasonableness of identifying Bel_0^{σ} (θ) with Bel_0 (θ/S) then this provides a justification for minimum cross entropy with respect to ME. The PV updates for RGEs in the discrete time domain can be found in this section. Reworking the outcomes of the previous theorem is the issue that the next theorem is trying to solve (Paris and Vencovska1992). by modifying them for the discrete case for all $q \in (-1, \infty)$ and the class of Rényi Generalized Entropies U_q^L . It is a fact that a directly analogous finding is satisfied by H_q^L with $q \in$ (-1,0), specifically with the notation of (Yager 2018), with N = H_r^L, $\mathrm{Bel}^{\sigma}(\theta) = U_r(K^{\sigma} + \mathrm{G})(\theta)$(see Theorem 3 of Paris and Vencovska1992), and directly comparable conclusion for $q = -1/2$ (named MTP in (previously discussed in, Paris and Vencovska1992, p.9) (as in Theorem 3 of Paris and Vencovska1992). Notably, the continuous time domain case for and the class of Rényi Generalized Entropies U_q^L was unable to solve the PV-updates beyond $q = -1/2$ (c.f., Paris and Vencovska1992, p.9). Therefore, the primary contribution of the following theorem is generalising the open interval research done by (Paris 1994; Paris and Vencovska 1992; Hawes 2007) on the open interval (-1,0).

Theorem $\mathrm{Bel}^{\sigma}(\theta) = H_q(K^{\sigma} + \mathrm{G})(\theta)$

To see this, we first observe, as in the case of $H_{-1/2}^L$ covered in (Paris and Vencovska1992),

$$\lim_{\varepsilon \searrow 0} \gamma^{\rightarrow}(\varepsilon) = \beta^{\rightarrow} \tag{3.58}$$

where $\beta^{\rightarrow} = H_q^L(K)$, the 2J –vector $\tau^{\rightarrow}(\varepsilon), \rho^{\rightarrow}(\varepsilon) = H_q^{L'}$((K+K$^{\sigma}$(S))+ (Bel(S) = ε) and $\beta^{\rightarrow}(\varepsilon) = \tau^{\rightarrow}(\varepsilon) + \rho^{\rightarrow}(\varepsilon)$.

Now, let:

$$\vartheta^{\rightarrow} = H_q^L(K^{\sigma} + \mathrm{G})(\theta) \tag{3.59}$$

and

$$H_q^L(\beta^{\rightarrow}(\varepsilon),\, y^{\rightarrow}, \varepsilon) = \sum_{i \notin I^L(K)}((\beta_i(\varepsilon) - \varepsilon y_i)^{q+1} \quad + (\varepsilon y_i)^{q+1})$$

$$= \sum_{i \notin I^L(K)}[(\beta_i(\varepsilon))^{q+1} \quad + (\varepsilon y_i)^{q+1} + \varepsilon(q+1) y_i(\beta_i - \varepsilon\theta y_i)^q] \tag{3.60}$$

for some $\theta = \theta(\beta^{\rightarrow}(\varepsilon),\, y^{\rightarrow}, \varepsilon), 0 \le \theta \le 1$.Suppose that $\frac{\rho^{\rightarrow}(\varepsilon)}{\varepsilon}$ had some subsequence $\frac{\rho^{\rightarrow}(\varepsilon_n)}{\varepsilon_n}$ converging to $\tau^{\rightarrow} \neq \vartheta^{\rightarrow}$.Then for some positive α,

$$\sum_{i \notin I^L(K)} \tau_i{}^{q+1} + \alpha < \sum_{i \notin I^L(K)} \vartheta_i{}^{q+1} \tag{3.61}$$

Therefore, we have for large n,

$$\sum_{i \notin I^L(K)} \vartheta_i^{q+1} > \sum_{i \notin I^L(K)} \left(\frac{\rho^{\rightarrow}(\varepsilon_n)}{\varepsilon_n}\right)^{q+1} + \frac{\alpha}{2} \tag{3.62}$$

By the open-mindedness of H_q^L, it follows that β_i for $i \notin I^L(K)$ are positive. Therefore, $\beta_i(\varepsilon_n)$ for $i \notin I^L(K)$ have a bound which is far from zero, hence it holds that for a fixed A and all n, $i \notin I^L(K), 0 \leq y_i \leq 1, (\beta_i(\varepsilon_n) - \varepsilon_n \theta y_i)^q < A$. Hence for n sufficiently large , we have

$$[H_q^L(\beta^{\rightarrow}(\varepsilon_n), \vartheta^{\rightarrow}, \varepsilon_n) - H_q^L(\beta^{\rightarrow}(\varepsilon_n), \frac{\rho^{\rightarrow}(\varepsilon_n)}{\varepsilon_n}, \varepsilon_n)]$$

$$\geq [\varepsilon_n{}^{q+1} \sum_{i \notin I^L(K)} \left(\vartheta_i^{q+1} - \left(\frac{\rho^{\rightarrow}(\varepsilon_n)}{\varepsilon_n}\right)^{q+1}\right) - 2(q+1) A\varepsilon_n / I^L(K) /]$$

$$\geq 0 \tag{3.63}$$

Since ($0 < q + 1 <1$), which contradicts the selection of $\beta^{\rightarrow}(\varepsilon_n)$,$\rho^{\rightarrow}(\varepsilon_n)$. The case H_q^L for $q = 1$ is already being investigated by (Paris 1994; Paris and Vencovska 1992). For all other positive values of q the query continues to be open.

3.6 Chapter Summary

This chapter focuses on the theory of uncertain reasoning and explores inference processes based on RGEs. It investigates the properties of these entropies in the discrete time domain and aims to define and analyze their behaviour. The chapter also discusses principles such as continuity, relativization, independence, and others, and seeks to update previous results to the class of RGEs in the discrete case . Additionally, it determines the PV-updates for this class.

4. Heavy-Tailed Stable $M/G/1$ Queue and Inductive Inferences Using Non-Extensive Maximum Entropy Formalisms

This chapter is devoted to the establishment of a new knowledge of information theoretic impact of the non-extensive parameter for investigating stable queueing systems. Consequently, a revealed influential impact has generated two novel state probabilities, namely the Rényian and Tsallisian closed form expressions solutions for the underlying stable $M/G/1$ queueing system. Additionally, a new underlying q-dependent families of underlying Rényian and Tsallisian service PDFs and cumulative distribution functions were derived and proven to make the newly derived solutions exact. More interestingly, it is shown that the information theoretic impact extends to the newly generated squared coefficients of variation in both Rényi and Tsallis cases. More potentially, we have proven that our derived solutions are credible by showing that they are reasoned by employing the consistency axioms. In mathematical terms, we proved that three axioms are satisfied and only one axiom was defied because of the non-extensivity impact.

4.1 Introduction

In both information theoretic and statistical physics terms, the provision of analytic inductive inference non-extensive ME-entropy (NME) closed form expressions for 'long-range' interactions physical systems of non-extensive information theoretic order were undertaken by (Rényi 1961; Tsallis 1988). Their work was an advancement of the information theoretic Shannonian "extensive" ME(EME) solutions depicting interactions of "short range” provided by (Shannon 1948).

The state probabilities $\{\mathrm{p(n)}$, n = 0,1,2, $\ldots\}$ of a stable $M/G/1$ queue were derived by maximising the EME functional, subject to two respective sets of mean value constraints, namely

Normalisation, $\sum_{\mathrm{n}} \mathrm{p(n)} = 1$ and Pollaczeck-Khinchin (P-K) mean queue length (MQL), $< \mathrm{n} > = \sum_{\mathrm{n}} \mathrm{np(n)}$ (Shore 1982) and $\rho = 1 - p(0)$ (4.1)

ρ serves as server utilization, (SU)(El-Affendi and Kouvatsos 1983).More importantly, for a stable $M/G/1$ queueing system ,EME closed form expression for the steady state probability along with both service time probability PDF and cumulative(CDF) distribution functions that make their provided analytic solution exact were devised by (El Affendi and Kouvatsos 1983), whose corresponding CDF had the following form:

$$F_s(t) = P(S \leq t) = 1 - \tau_s exp(-\mu t \, \tau_s) \tag{4.2}$$

$$\tau_s = \frac{2}{1+C_S^2} \tag{4.3}$$

μ serves as the mean service rate and C_s^2 denotes the squared coefficient of variation, (SCV) of the service times. For more details see (El-Affendi and Kouvatsos 1983; Kouvatsos 1986; Kouvatsos 1988)). The evaluation of long-range interactions’ Tsallisian closed form solution based on the axioms of consistency as well as the applicability of Shannonian /Tsallisian ME solutions in analysing several heavy-tailed bursty queues were developed in (Kouvatsos 2010;Kouvatsos 2011).

4.2 General Systems and Inductive ME Formalisms

4.2.1 ‘Classical’ Shannon’s EME Formalism with ‘Short-Range’ Interactions

The EME Shannonian entropic functional(Shannon 1948), $\mathrm{H}_{1,S}(\mathrm{p})$ for any general physical system Q depicting short range interactions, is given by:

$$H_{1,S}(p) = -c \sum_{S_n \in S} p_{1,S}(S_n) \log p_{1,S}(S_n) \tag{4.4}$$

with constant c> 0 and $\{p_{1,S}(S_n),\ S_n \in S = \{S_n, n = 0, 1, 2, \dots\}$ to denote EME state probabilities for finite or countably infinite set S of configurations or states $\{S_n, n = 0, 1, 2, \dots\}$.

The short-range interactions have potential interpretations in Statistical Physics(Statistics 2007). More interestingly, traffic flows in queues with short-range dependence (SRD) have been explored in (Kleinrock 1976).

4.2.2 Rényi's and Tsallis's NME Formalisms with Long-Range Interactions

For a general system Q with an integer number of possible (microscopic) configurations or states N (> 0) and "long-range interactions," such as gravity in statistical physics (see Statistics 2007), energy and entropy are no longer significant quantities. This makes the physical system Q more complex because the state probability distribution linked to, say, energy can no longer be predicted by maximising the Shannon's extensive entropy, $H_{1,S}(p)$, due to its heavy tails and power law behaviour.

To this end, for the afore-mentioned physical system Q with a finite or countably infinite set S of configurations or states $\{S_n,\ n = 0, 1, 2, \dots\}$ and 'long-range interactions' proposed by (Tsallis 1988) and (Rényi 1961) NME functionals are defined by

$$H_{q,T}(p) = \frac{c}{q-1}\left\{1 - \sum_{S_n \in S} p_{q,T}(S_n)^q\right\} \tag{4.5}$$

$$H_{q,R}(p) = \frac{c}{1-q}\ln\left\{\sum_{S_n \in S} p_{q,R}(S_n)^q\right\} \tag{4.6}$$

respectively, for any constant $c > 0$.

4.3 EME Consistency Axioms coined with EME formalisms

The undertaken method (Shore 1982; Shore 1980)) was built on the main hypothesis : "When there are several techniques to analyse a physical system Q that take into account the same prior knowledge limitations, using the EME principle should result in consistent state probabilities for the system". For a more detailed account on the physical interpretation of these axioms (see Shore 1980; Statistics 2007).

In a broader sense, we have explored the credibility of both derived Rényian and Tsallisian NME formalisms by employing these axioms of consistency. See appendices chapter for more details. Fundamentally, this strongly supports the suitability of these novel derivations to quantitively investigate heavy-tailed queue dynamic systems with long range interaction.

4.4. Long-Range Interactions' Stable M/G/1 Queue

This section develops new analytical solutions for the steady state probabilities of a stable M/G/1 queueing system according to the normalisation, SU, and MQL criteria using Rényi and Tsallis' generalised NME formalisms.

The notation used throughout the analysis is as follows:

- $X = S\ (Shannon)\ with\ q = 1, T$ (Tsallis) $and\ R$ (Rényi) $with\ 0.5 < q < 1$;
- $p_{q,X}(n)$, steady state probability of having n jobs in the $M/G/1\ queue, n = 0, 1, 2, \ldots$;
- λ, mean arrival rate; μ, mean service rate; $\rho = \lambda/\mu, SU; < n > = \sum_{n=0}^{\infty} np_{q,X}(n), MQL$;
- $C_{s,q,X}^2, SCV$ of the $GE -$ type service (s) time distribution, $GE_{q,X}$;
- $f_{s,q,X}(t)\ and\ F_{s,q,X}(t)$, Probability density and cumulative functions of the service

  time distribution;
- $F_{s,q,X}^*(\theta)$, Laplace transform of $f_{s,q,X}(t)$.

4.4.1 Foundation notation: The Shannonian EME State Probability of Stable $M/G/1$ Queue

The $M/G/1$ queue's EME steady state probability that maximises Shannon entropy(El-Affendi and Kouvatsos 1983)

$$H(p_{1,S}) = -\sum_{n=0}^{\infty} p_{1,S}(n) \ln(p_{1,S}) \tag{4.7}$$

Under prior information and mean value constraints:

- $\sum_{n=0}^{\infty} p_{1,S}(n) = 1$ (4.8)
- SU,

$$p_{1,S}(0) = \sum_{n=0}^{\infty} h(n) p_{1,S}(n) = 1 - \rho,\ \rho = \frac{\lambda}{\mu} \tag{4.9}$$

where

$$h(n) = \begin{cases} 1 & n = 0 \\ 0 & n = 1,2,\ldots \end{cases} \tag{4.10}$$

- P-K MQL,

$$< n > = \sum_{n=0}^{\infty} n p_{1,S}(n) = \frac{\rho}{2}\left(1 + \frac{1+\rho C_{S,1,S}^2}{1-\rho}\right) \tag{4.11}$$

is given by

$$p_{1,S}(n) = \begin{cases} p_{1,S}(0), & n = 0 \\ p_{1,S}(0)\tau_S x^n & n > 0 \end{cases} \tag{4.12}$$

where $p_{1,S}(0) = 1 - \rho$, $\tau_s = 2/(1+ C_{S,1,S}^2)$ and $x = \frac{<n> - \rho}{<n>}$.

4.4.2 The Stable $M/G/1$ Queue's Rényian and Tsallisian NME Steady State Probabilities

Theorem 4.1

The Rényian and Tsallisian NME steady-state probabilities of stable $M/G/1$ queueing system, $p_{q,X}(n),\ n = 0,1,2,\ldots$,$\mathrm{X} = \mathrm{R}, \mathrm{T}$, respectively, subject to (4.8), (4.9) and (4.11) constraints are determined by

$$p_{q,R}(n) = \begin{cases} p_{q,R}(0) & n = 0 \\ p_{q,R}(0)\tau_S^{\frac{1}{q}} x^n & n > 0 \end{cases} \tag{4.13}$$

$$p_{q,T}(n) = \begin{cases} p_{q,T}(0) & n = 0 \\ p_{q,T}(0)\tau_s{}^q x^n & n > 0 \end{cases} \tag{4.14}$$

Such that

$$p_{q,R}(0) = p_{q,T}(0) = 1 - \rho \tag{4.15}$$

Here τ_s and *x* serve as Lagrange's multipliers under (4.9) and (4.11) constraints are devised by

$$\tau_s = \frac{2}{1+ C_{s,1,S}^2} \tag{4.16}$$

$$x = \begin{cases} \dfrac{\rho}{\left(\rho+(1-\rho)\left(\frac{2}{1+C_{s,1,S}^2}\right)^{\frac{1}{q}}\right)}, & \text{Rényi} \\ \dfrac{\rho}{\left(\rho+(1-\rho)\left(\frac{2}{1+C_{s,1,S}^2}\right)^{q}\right)}, & \text{Tsallis} \end{cases} \tag{4.17}$$

with

$$\frac{\rho(1-x)}{(1-\rho)x} = \begin{cases} \tau_s{}^{\frac{1}{q}}, & \text{Rényi} \\ \tau_s{}^q, & \text{Tsallis} \end{cases} \tag{4.18}$$

Proof

The maximization of Rényi's entropy under constraints (4.8), (4.9) and (4.11) puts the Lagrangian into the following form:

$$[\frac{q}{(1-q)\left(\sum_{n=0}^{\infty}\left(p_{q,R}(n)\right)^q\right)}\sum_{n=0}^{\infty}\left(p_{q,R}(n)\right)^{q-1} - \alpha(\sum_{n=0}^{\infty} h(n)) - \beta(\sum_{n=0}^{\infty} 1) - \gamma(\sum_{n=0}^{\infty} n)] = 0 \tag{4.19}$$

Hence, $p_{q,R}(n)$ reads as follows:

$$p_{q,R}(n) = a(1 + b\,\frac{(1-q)}{q}\,n)^{\frac{1}{q-1}} \tag{4.20}$$

Satisfying that

$$a = \left(\frac{(1-q)\left(\sum_{n=0}^{\infty}\left(p_{q,R}(n)\right)^{q}\right)}{q}\right)(\,\alpha h(n)+\beta))^{\frac{1}{q-1}},\;\; b = \frac{\gamma}{(\,\alpha h(n)+\beta)}\frac{(1-q)}{q} \qquad (4.21)$$

Equ.(4.20) implies:

$$a = p_{q,R}(0) \qquad (4.22)$$

Therefore, $p_{q,R}(n)$ is:

$$p_{q,R}(n) = p(0)\left(1+b\frac{(1-q)}{q}n\right)^{\frac{1}{q-1}} \;\; ,b\; > \;0 \text{ and } q > 0, q \neq 1 \qquad (4.23)$$

We can re-write the analytic formula (4.23) by replacing the expression

$(1+b\frac{(1-q)}{q}n)^{\frac{1}{q-1}}$ with its equivalent expression, ${\tau_s}^{\frac{1}{q}}x^n$. We arrive at the formula:

$$p_{q,R}(n) = \quad p_{q,R}(0){\tau_s}^{\frac{1}{q}}x^n \qquad (4.24)$$

where

$${\tau_s}^{\frac{1}{q}}x^n \;=\; \left(1+b\frac{(1-q)}{q}n\right)^{\frac{1}{q-1}} \qquad (4.25)$$

Combining constraint (4.8)with the derived formula (4.21), together with $M/G/1$ queue analysis that

$$1 = \; p(0)+p(0){\tau_s}^{\frac{1}{q}}\frac{x}{(1-x)} \Longrightarrow {\tau_s}^{\frac{1}{q}} = \frac{\rho(1-x)}{(1-\rho)x}\,,\; \tau_s = \frac{2}{(1+y_s\,)}\,,\; y_s = \frac{(C_{s,q,R}^2-1)}{2} \qquad (4.26)$$

Moreover, $\tau_s \;= \frac{2}{1+\,C_{s,q,R}^2}$ and ${\tau_s}^{\frac{1}{q}} = \frac{\rho(1-x)}{(1-\rho)x}$, which clearly implies $x = \frac{\rho}{\left(\rho+(1-\rho)\left(\frac{2}{1+C_{S,1,S}^2}\right)^{\frac{1}{q}}\right)}$.

Focusing on the maximisation of Tsallis's NME functional, subject to constraints (4.8), (4.9) and (4.11), the corresponding Lagrangian equation is

$$[q\sum_{n=0}^{\infty}\left(p_{q,T}(n)\right)^{q-1} - \alpha'(\sum_{n=0}^{\infty}h(n))) - \beta'(\sum_{n=0}^{\infty}1) - \gamma'(\sum_{n=0}^{\infty}n)] = 0 \quad (4.27)$$

where $p_{q,T}(n)$ is in the form

$$p_{q,T}(n) = a'(1 + w(1-q)n)^{\frac{1}{q-1}} \quad (4.28)$$

and

$$a' = [\left(\frac{1}{q}\right)(\alpha' h(n) + \beta')]^{\frac{1}{q-1}}, w = \frac{\gamma'}{(\alpha' h(n)+\beta')}(1-q) \quad (4.29)$$

From (4.28) – (4.29), we obtain that:

$$a' = p_{q,T}(0) \quad (4.30)$$

Hence, it is implied that:

$$p_{q,T}(n) = p_{q,T}(0)(1 + wq(1-q)n)^{\frac{1}{q-1}}\, b > 0\,, q > 0, q \neq 1 \quad (4.31)$$

Hence, we can re-write (4.31) in the form

$$p_{q,T}(n) = \quad p_{q,T}(0){\tau_s}^q x^n \quad (4.32)$$

where

$${\tau_s}^q x^n = \ (1 + wq(1-q)n)^{\frac{1}{q-1}} \quad (4.33)$$

Engaging (4.8) and (4.21), we have

$$1 = \ p_{q,T}(0) + p_{q,T}(0){\tau_s}^q \frac{x}{(1-x)}$$

Hence,

$${\tau_s}^q = \frac{\rho(1-x)}{(1-\rho)x} \quad (4.34)$$

Moreover, it is clearly follows that:

$$x = \frac{\rho}{\left(\rho+(1-\rho)\left(\frac{2}{1+C_{S,1,S}^2}\right)^q\right)} \tag{4.35}$$

4.4.3. Closed form expressions of Rényian and Tsallisian time probability density functions

This section deals with the derivation of both Rényian and Tsallisian service time PDFs that make $\{p_{q,R}(n), p_{q,T}(n), n = 0, 1,2, . .\}$, respectively,(see (4.13) and (4.14)) exact.

4.4.3.1 $\boldsymbol{f_{s,q,R}(t)}$ *and* $\boldsymbol{f_{s,q,T}(t)}$

Theorem 4.2

If given $p_{q,R}(n), p_{q,T}(n)$ of (4.13) and (4.15) respectively, are exact whenever the corresponding closed form expressions $f_{s,q,R}(t)$ *and* $f_{s,q,T}(t)$ are determined by

$$f_{s,q,R}(t) = \left(1 - \tau_s{}^{\frac{1}{q}}\right)u_0(t) + \mu\tau_s{}^{\frac{2}{q}}e^{-\mu t\tau_s{}^{\frac{1}{q}}} \tag{4.36}$$

$$f_{s,q,T}(t) = (1 - \tau_s{}^q)u_0(t) + \mu\tau_s{}^{2q}e^{-\mu t\tau_s{}^q} \tag{4.37}$$

$u_0(t)$ serves as

$$\mathrm{u}_0(\mathrm{t}) = \begin{cases}\infty, & \mathrm{t} = 0\\ 0, & \mathrm{t} \neq 0\end{cases}, \int_{-\infty}^{\infty} \mathrm{u}_0(\mathrm{t}) = 1 \text{ and } \quad \tau_\mathrm{s} = \frac{2}{1+\mathrm{C}_{\mathrm{s},1,\mathrm{S}}^2} \tag{4.38}$$

Proof

Following (Kleinrock 1976), we can write the Rényian z-transform $Q_{q,R}(z)$ of $p_{q,R}(n)$ of (4.13) is

$$Q_{q,R}(z) = \sum_{n=0}^{\infty} p_{q,R}(n)z^n, \ |\mathrm{z}| < 1 \tag{4.39}$$

Hence, by replacing $p_{q,R}(n)$ of (4.13) into (4.39), it follows that :

$$Q_{q,R}(z) = \sum_{n=0}^{\infty} p_{q,R}(n) z^n = p_{q,R}(0) + \frac{p_{q,R}(0)\tau_s^{\frac{1}{q}} xz}{(1-xz)} \quad = \frac{p_{q,R}(0)(1-xz\left(1-\tau_s^{\frac{1}{q}}\right))}{1-xz} \tag{4.40}$$

Engaging P/K transformation (Kleinrock 1976), we get

$$Q_{q,R}(z) = \frac{p_{q,R}(0)(1-z)\left(F_{s,q,R}^*(\lambda-\lambda z)\right)}{F_{s,q,R}^*(\lambda-\lambda z)-z} \tag{4.41}$$

where

$$F_{s,q,R}^*(\theta) = E\left[e^{-\theta s}\right] = \int_0^{\infty} e^{-\theta t} f_{s,q,R}(t) dt \tag{4.42}$$

is the $f_{s,q,R}(t)$'s Laplace-Stieltjes transform. Clearly, $Q_{q,R}(0) = p_{q,R}(0)$ and $Q_{q,R}(1) = 1$.

Engaging (4.40) and (4.41), we have

$$F_{s,q,R}^*(\lambda - \lambda z) = \frac{\mu\tau_s^{\frac{1}{q}} + (\lambda-\lambda z)(1-\tau_s)^{\frac{1}{q}}}{\mu\tau_s^{\frac{1}{q}} + (\lambda-\lambda z)} \tag{4.43}$$

Substituting θ for $(\lambda - \lambda z)$, the expression (4.43) becomes:

$$F_{s,q,R}^*(\theta) = \frac{\mu\tau_s^{\frac{1}{q}} + \theta(1-\tau_s^{\frac{1}{q}})}{\mu(\tau_s)^{\frac{1}{q}} + \theta} \tag{4.44}$$

By inverting Laplace-StieltjesTransform, $F_{s,q,R}^*(\theta)$ the $GE_{q,R}$–type PDF $f_{s,q,R}(t)$(see (4.40)) of Theorem 4.2 is devised.

In a similar analogy to the logical consequence of the undertaken proof in the Rényian case, the NME solution $p_{q,T}(n), n = 0,1,2,...$ has the z-transform $Q(z)$,

$$Q_{q,T}(z) = \sum_{n=0}^{\infty} p_{q,T}(n) z^n = \frac{p_{q,T}(0)(1-xz(1-\tau_s{}^q))}{1-xz} \tag{4.45}$$

Using the P/K transform equation

$$Q_{q,T}(z) = \frac{p_{q,T}(0)(1-z)\left(F_{s,q,T}^*(\lambda-\lambda z)\right)}{F_{s,q,T}^*(\lambda-\lambda z)-z} \tag{4.46}$$

where $F^{*}_{s,q,T}(\theta) = E\left[e^{-\theta s}\right] = \int_0^{\infty} e^{-\theta t} f_{s,q,T}(t)dt$ (4.47)

serves as the of the service time distribution PDF's Laplace-Stieltjes transform. We have

$$Q_{q,T}(0) = p_{q,T}(0) \; and \; Q_{q,T}(1) = 1$$

By (4.45) and (4.46), we have

$$F^{*}_{s,q,T}(\lambda - \lambda z) = \frac{\mu(\tau_s)^q + (\lambda-\lambda z)(1-(\tau_s)^q)}{\mu(\tau_s)^q + (\lambda-\lambda z)} \quad (4.48)$$

by replacing θ for $(\lambda - \lambda z)$, (4.58) re-writes to

$$F^{*}_{s,q,T}(\theta) = \frac{\mu {\tau_s}^q + \theta(1-{\tau_s}^q)}{\mu(\tau_s)^q + \theta} \quad (4.49)$$

The inversion of Laplace-Stieltjes transform implies that PDF $f_{s,q,T}(t)$(see (4.37)) of Theorem 4.2 is devised.

The following Corollary 4.2.1 deals with the characterisation of the CDFs, $\{F_{s,q.T}(\mathrm{t}), F_{s,q.T}(\mathrm{t})\}$.

Corollary 4.2.1

The CDFs of the service times *GE*$_{q,R}$ and *GE*$_{q,T}$ types PDFs (c.f., (4.36) and (4.37)) are respectively determined by

$$F_{s,q,R}(t) = 1 - {\tau_s}^{\frac{1}{q}} e^{-\mu {\tau_s}^{\frac{1}{q}} t} \quad (4.50)$$

$$F_{s,q,T}(t) = 1 - {\tau_s}^q e^{-\mu {\tau_s}^q t} \quad (4.51)$$

where $\tau_s = 2/(1 + C^2_{s,1,S})$.

Proof

We have

$F_{s,q,R}(t)$= $\int_0^t f_{s,q,R}(x)\,dx$= $\int_0^t \left(1-{\tau_s}^{\frac{1}{q}}\right) u_0(x)\,dx + \mu {\tau_s}^{\frac{2}{q}} \int_0^t e^{-\mu {\tau_s}^{\frac{1}{q}} x}\,dx$

$$= \left(1-{\tau_s}^{\frac{1}{q}}\right) + \frac{\mu {\tau_s}^{\frac{2}{q}}}{\mu {\tau_s}^{\frac{1}{q}}}\left(1-e^{-\mu {\tau_s}^{\frac{1}{q}} t}\right)$$

$$= \left(1-{\tau_s}^{\frac{1}{q}}\right) + {\tau_s}^{\frac{1}{q}}\left(1-e^{-\mu {\tau_s}^{\frac{1}{q}} t}\right)$$

$= 1-{\tau_s}^{\frac{1}{q}} e^{-\mu {\tau_s}^{\frac{1}{q}} t}$ (See (4.50))

Moreover, Tsallis's service time CDF, $F_{s,q,T}(t)$ is characterised as follows:

$F_{s,q,T}(t) = \int_0^t f_{s,q,T}(x)\,dx = \int_0^t (1-{\tau_s}^q) u_0(x)\,dx + \mu {\tau_s}^{2q} \int_0^t e^{-\mu {\tau_s}^q x}\,dx$

$$= (1-{\tau_s}^q) + \frac{\mu {\tau_s}^{2q}}{\mu {\tau_s}^q}\left(1-e^{-\mu {\tau_s}^q t}\right)$$

$$= (1-{\tau_s}^q) + {\tau_s}^q\left(1-e^{-\mu {\tau_s}^q t}\right)$$

$= 1-{\tau_s}^q e^{-\mu {\tau_s}^q t}$ (See (4.51))

It is to be noted that as $q \to 1$, both derived CDFs of (4.50) and (4.51) tend to the limiting Shannonian CDF case of (El-Affendi and Kouvatsos 1983), $F_{s,q,T}(t) = 1 - \tau_s e^{-\tau_s \mu t}$ with $\tau_s = \frac{2}{C_{s,1,S}^2+1}$.

Corollary 4.2.2

The SCVs of the service time corresponding to service time PDFs of (4.42) and (4.43) are given by:

$$E\left(s_{q,R}\right) = \frac{1}{\mu} \quad (4.52)$$

$$E\left(s_{q,R}^2\right) = \frac{2}{\mu^2 {\tau_s}^{\frac{1}{q}}} \quad (4.53)$$

$$C^2_{s,q,R} = \frac{E(s^2_{q,R})}{(E(S_{q,R}))^2} - 1 = \frac{\left(2 - \tau_s^{\frac{1}{q}}\right)}{\tau_s^{\frac{1}{q}}} \tag{4.54}$$

$$E(S_{q,T}) = \frac{1}{\mu} \tag{4.55}$$

$$E(s^2_{q,T}) = \frac{2}{\mu^2 \tau_s{}^q} \tag{4.56}$$

$$C^2_{s,q,T} = \frac{E(S^2_{q,T})}{\left(E(S_{q,T})\right)^2} - 1 = \frac{(2 - \tau_s{}^q)}{\tau_s{}^q} \tag{4.57}$$

where τ_s = 2/(1+ $C^2_{s,1,S}$).

Proof

We have

$$E(S_{q,R}) = \int_0^\infty t\, f_s(t)dt = \int_0^\infty t\mu\tau_s^{\frac{2}{q}}\, e^{-\mu\tau_s^{\frac{1}{q}}t}dt = \mu\tau_s^{\frac{2}{q}} \int_0^\infty t\, e^{-\mu\tau_s^{\frac{1}{q}}t}dt \tag{4.58}$$

The gamma function, $\Gamma(m)$ is given by

$$\Gamma(m) = \int_0^\infty w^{m-1}\, e^{-w}dw \tag{4.59}$$

Substituting $w = \mu\tau_s^{\frac{1}{q}}t$ on (4.59), we have, $E(\mathrm{s_R}) = \frac{\mu\tau_s^{\frac{2}{q}}}{\mu^2\tau_s^{\frac{2}{q}}}\,\Gamma(2) = \frac{1}{\mu}$ (because $\Gamma(2) = 1$)

Moreover, it can be seen that:

$$E(s^2_{q,R}) = \int_0^\infty t^2\, f_s(t)dt = \int_0^\infty t^2\mu\tau_s^{\frac{2}{q}}\, e^{-\mu\tau_s^{\frac{1}{q}}t}dt = \mu(\tau_s)^{\frac{2}{q}} \int_0^\infty t^2\, e^{-\mu\tau_s^{\frac{1}{q}}t}dt \tag{4.60}$$

and setting $\mathrm{w} = \mu\tau_s^{\frac{1}{q}}t$, $E(s^2_{q,R})$ is given by (4.53) and subsequently, $C^2_{s,q,R}$ by (4.54).

Engaging an analogous approach, (4.55) – (4.57) for Tsallis's case can be obtained.

4.5. Two Case Studies with explanations

This section gives two case studies that serve as examples of how $p_{q,T}(n)$ and $p_{q,R}(n)$ $(\text{c.f.}, (4.13)\text{and } (4.14))$ can be used numerically to two different areas to get conclusions about a stable heavy-tailed $M/G/1$ queueing system.

- **Case Study 1**

An exploration of the queue tails of $\mathrm{p_{q,R}(n)}$ and $\mathrm{p_{q,T}(n)}$(c.f., (4.13) and (4.14) in connection with the impact of non – extensive order q, which causes a depicted long-range interaction is undertaken in case study 1.

In principle, it performs numerical tests in this context and plots state probabilities vs n = 1, 2, with i) μ= 0.75 ii)μ = 1.00 (i.e., SU, =75%). Selected original $\mathrm{C}^2_{\mathrm{s,1,S}}$= 2, 5, 10, 20, and 50, were employed in (4.15) (such that $q \to 1$) and order q "long-range" interactions of order q (c.f., Figures (4.1)- (4.10)).

The queue tails of $p_{q,R}(n)$ and $p_{q,T}(n)$, (c.f., (4.13) and (4.14)) are investigated in Case Study 1 in relation to the impact of q, n = 1, 2,...In principle, it performs numerical tests in this context and plots state probabilities vs n = 1, 2,... with i)μ= 0.75 ii)μ = 1.00 (i.e., SU, =75%) combined with $\mathrm{C}^2_{\mathrm{s,1,S}}$= 2, 5, 10, 20, and 50, were employed in constraint (4.11)(with q=1) and with an order q "long-range" interactions (c.f., Figures (4.1)-(4.6)).

Potentially, heavy queue tails are captured by both Rényian and Tsallisian formalisms(c.f., (4.13) and (4.14)) because of the significant impact of the non-extensive exponent q (c.f., Figures (4.1)-(4.5)).

Clearly, this manifests the influential impact of the information theoretic non-extensivity that is entwined with the provision of longer ranges of interactions to cause the heavy queue tails phenomena by both Rényian and Tsallisian formalisms.

More interestingly, it was clear that the queue tails in the Rényian case were heavier than that of Tsallisian. Additionally, by the progressive increase of $C^2_{s,1,S}$, these depicted heavy tails are flattened progressively until they take a linearity phase(straight lines, corresponding to $C^2_{s,1,S} = 50$). It is to be noted that the progressive increase of q impacts the Rényian formalism to exhibit heavier queue tails, whereas q → 1.00, the Tsallisian case depicts heavier queue tails by the increase of q.

Both $\mathrm{p}_{\mathrm{q,R}}(\mathrm{n}), \mathrm{p}_{\mathrm{q,T}}(\mathrm{n})$ can be seen to capture large queue tails (cf., Figures (4.1)-(4.5)) as a direct consequence of the non-extensivity of the order q, which causes 'long-range' interactions. Tsallis' state probability $p_{q,T}(n),\ n = 1,2, \ldots$, on the other hand, captures lighter queue tails than Rényi's state probability $p_{q,R}(n),\ n = 1,2, \ldots,\ C^2_{s,1,S}$ = 2, 5, 10, 20, and 50 for all combined values of both SCVs, and information order q = 0.55, 0.70, 0.85, and 1.00, as needed. By the progressive rise of $C^2_{s,1,S}$, the heavy tails become "flatter," and the corresponding curves transform into straight lines (linearity phase) (c.f., $C^2_{s,1,S} = 50$).

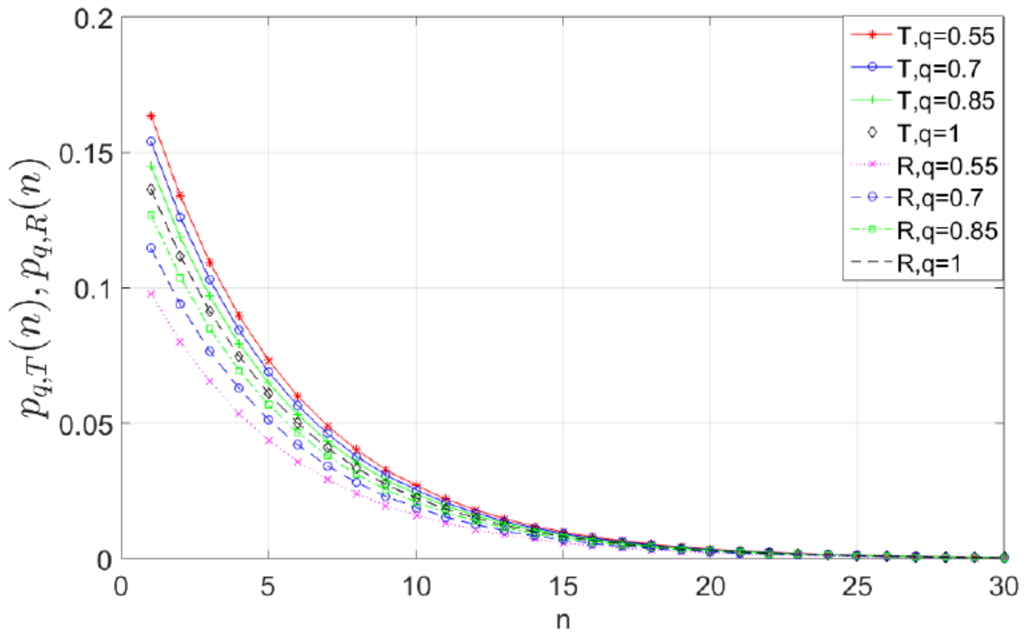


Figure 4.1: $Tsallisian\ \ and\ Rényian\ formalisms$ vs n for $C^2_{s,1,S}$ = 2

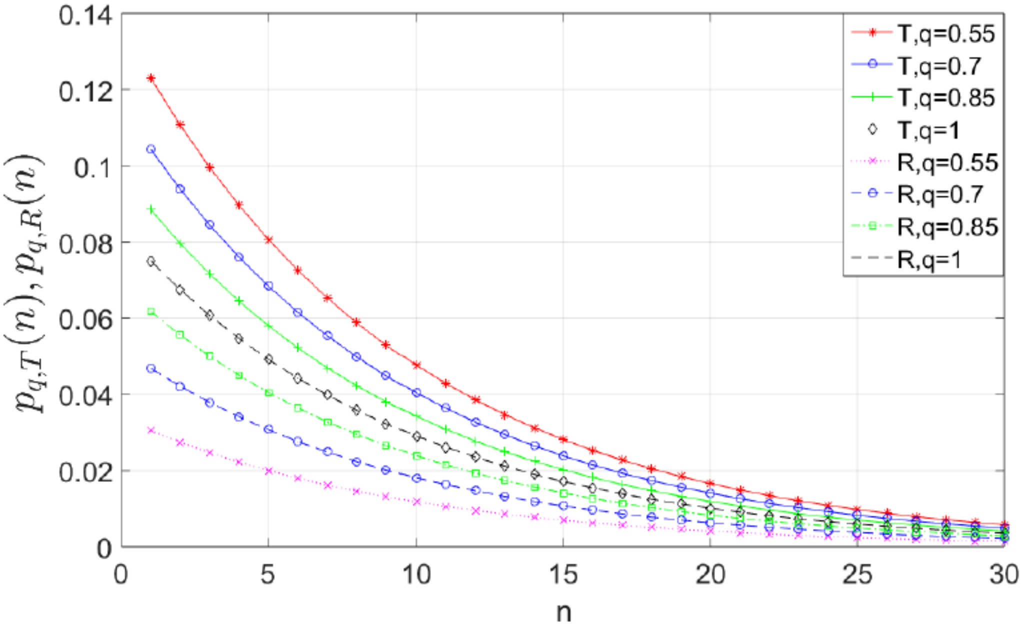


Figure 4.2: Tsallisian and Rényian formalisms vs n for $C_{s,1,S}^2 = 5$

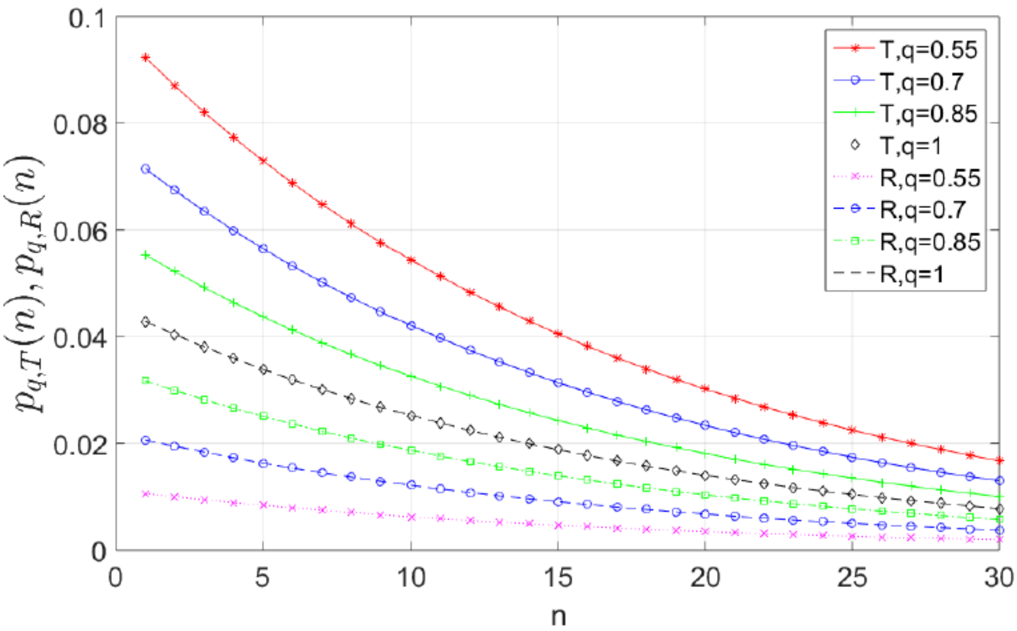


Figure 4.3: $Tsallisian\ \ and\ Rényian\ formalisms$ vs n for $C_{s,1,S}^2 = 10$

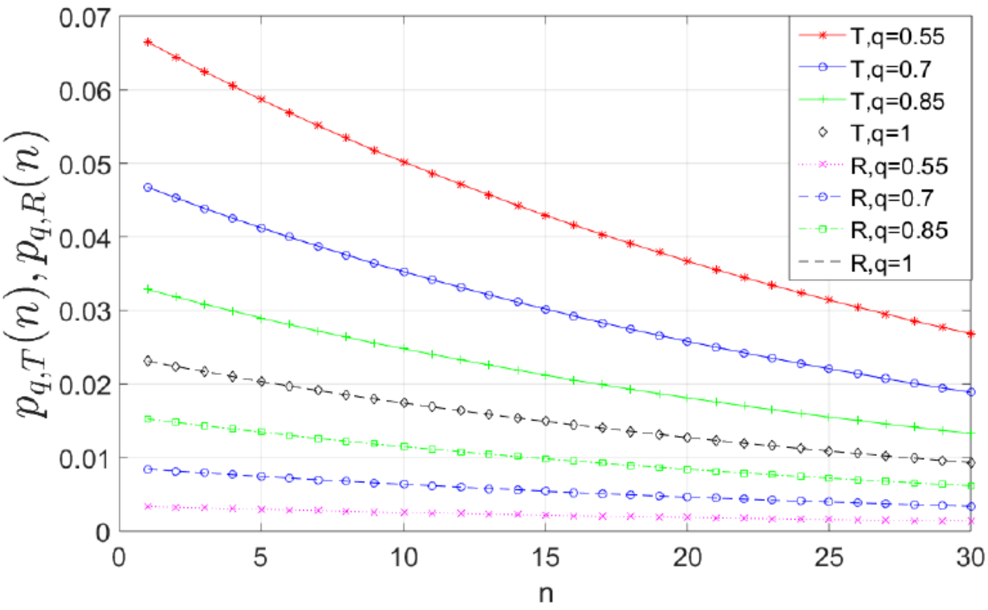


Figure 4.4: $Tsallisian\ \ and\ Rényian\ formalisms$ vs n for $C_{s,1,S}^{2}$ = 20

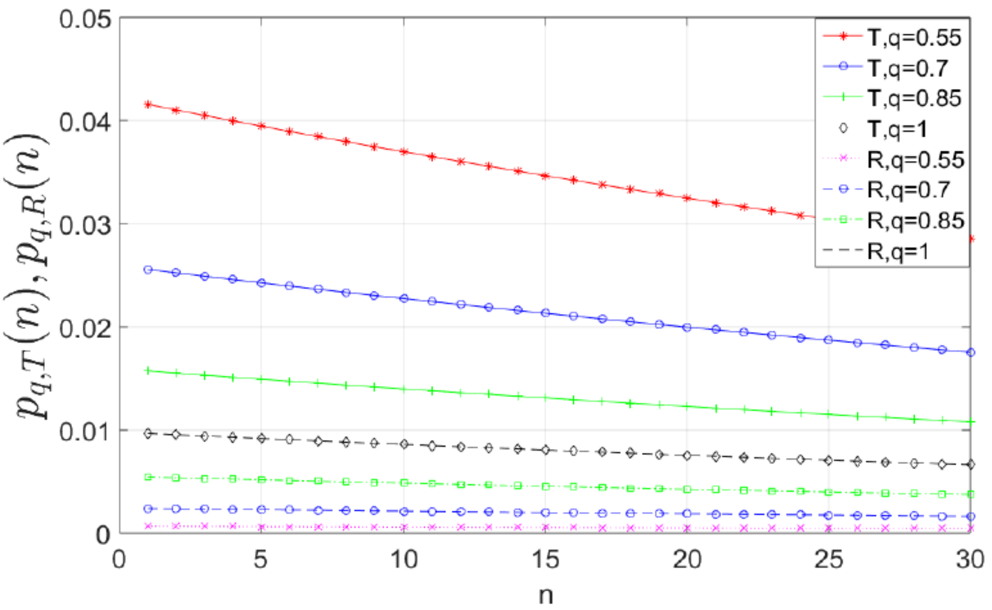


Figure 4.5: $Tsallisian\ \ and\ Rényian\ formalisms$ vs n for $C_{s,1,S}^{2}$ = 50

- **Case Study 2**

In principle, this case study investigates how q impacts the behaviour of $C_{s,q,R}^{2}$ and $C_{s,q,T}^{2}$ (c.f., (4.54), (4.57)) for our chosen values of $C_{s,1,S}^{2}$.

The curves of q vs Rényian and Tsallisian SCVs , with q = 0.55, 0.70, 0.85, and 1.00 for each SCV, and $C^2_{s,1,S}$ = 2, 5, 10, are drawn because of the numerical experiments shown in Figure (4.6). The limiting case of the SCVs $C^2_{s,q,R}$ and $C^2_{s,q,T}$vs q in Figure (4.6) is convergent which agrees with our expectations at $C^2_{s,1,S}$ (as $q \to 1$). This holds by the progressive increase of both $C^2_{s,1,S}$ and q. Clearly, the lower variability of $C^2_{s,q,T}$ compared to $C^2_{s,q,R}$ can be seen from Figure (4.6), which illustrates that $C^2_{s,q,T} < C^2_{s,q,R}$ as q increases.

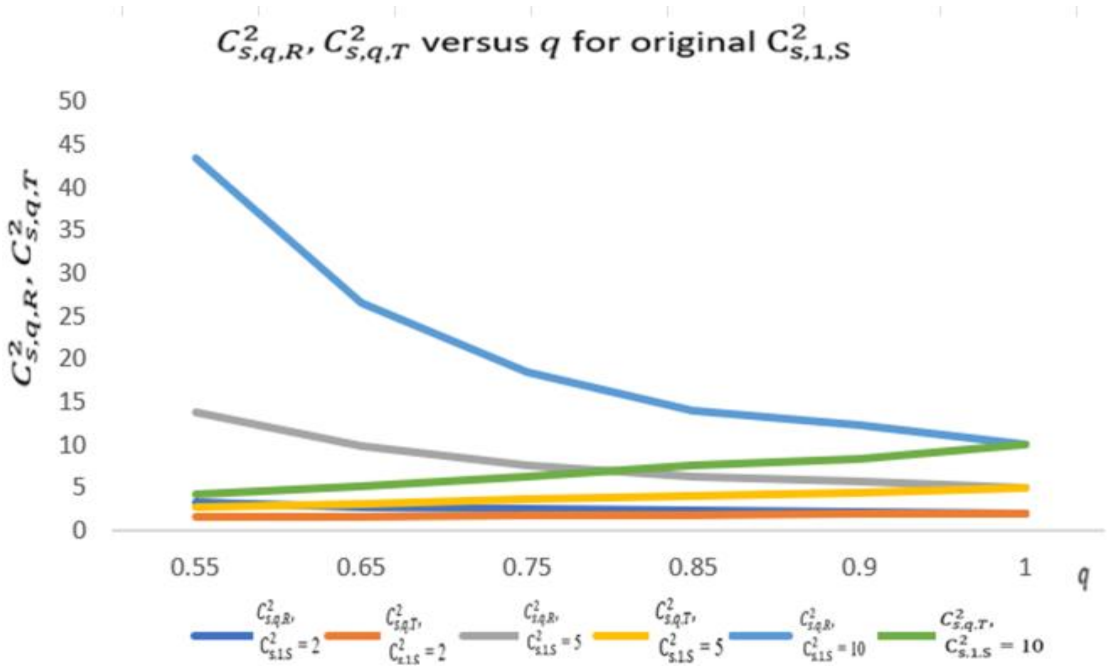


Figure 4.6 : q vs Rényian and Tsallisian SCVs for original $C^2_{s,1,S}$ = 2,5,10

4.6 Chapter Summary

This chapter is pivotal to the establishment of contemporary information theoretic queueing theory. Briefly, it contributes to revealing that the Rényian and Tsallisian formalisms of the underlying stable $M/G/1$ queueing system are characterized by a of q-dependent families, which reduce to the only available formalism in the literature, namely the Shannonian formalism.

More intriguingly, we have derived both the service time distribution and cumulative functions that make these formalisms exact. We have also extended the undertaken research to the provision of new-to the knowledge Rényian and Tsallisian squared coefficients of variation. The credibility of our formalisms had been proven by using the four well-known consistency axioms.

Future research pathways include the exploration of more new formalisms for the stable $M/G/1$ queueing systems and other available queueing systems in the literature, by employing other generalized entropies.

5. The Influential Role of Information Geometry on The Analysis of The Stable $M/G/1$ Queue Manifold

This chapter unifies queueing theory with several mathematical fields, including differential geometry, information theory, matrix theory, and information geometry (IG). More importantly, the application of info-geometric techniques to queueing theory is a powerful approach to studying the stability of queues and allowing to revolutionize classical queueing theory by employing innovative IG techniques.

5.1 Introduction

The appeal of information geometry (IG) lies in the way differential geometry (DG) is used to describe the structure of statistical models. Fundamentally, based on the IG, we are stepping ahead to a new era of Info-Geometric Queueing Theory (IGQT).
Numerous study areas, including statistical inference, stochastic control, and neural networks, make extensive use of information geometry (IG).

Geodesics are representations of straight lines in Euclidean space and resemble them in many ways. Following the celebrated theory of Einsteinian General Relativity (GR), Geodesic objects move through curved space-time, which causes the ideal time between two points to be extremely long. Thus, the geometry of curved space and the geometry of space-time are both explained by the same mathematics. A geodesic line is a "straight line on a curved surface" that minimises the distance between two points.

5.2 Main Definitions in Information Geometry

Definition 5.2.1 $n-\mathbf{dimensional\ distribution\ manifold}$ (Li et al. 2007)

1.If x specifies a random variable in a sample space X with a probability density function $p(x,\theta)$, then we define an n –dimensional distribution manifold M to be written as $M = \{lnp(x,\theta)| = (\theta_1, \theta_2, .., \theta_n) \in \mathbb{R}^n\}$. It is to be noted that $\theta = (\theta_1, \theta_2, .., \theta_n)$ are called system coordinates.

Definition 5.2.2 Potential Function (Li et al 2007)

The potential function $\Psi(\theta)$ of definition (5.2.1) is defined to be the standalone part of coordinates alone $of\ (-\ln p(x;\ \theta))$ that only contains coordinates.

Definition 5.2.3 The Fisher's metric(FM), or Fisher Information matrix (FIM) and α-Connection (Dodson 2005)

1. For $\Psi(\theta)$ (c.f., definition 5.2.2),FIM is the nxn matrix devised by

$$[g_{ij}] = \left[\frac{\partial^2}{\partial\theta^i \partial\theta^j}\left(\Psi(\theta)\right)\right], \quad i,j = 1,2,3, ... \tag{5.1}$$

We are here performing partial differentiation with respect to θ coordinates.

2. The inverse matrix of $[g_{ij}]$ of (5.1) is defined by

$$[g^{ij}] = ([g_{\ ij}])\ )^{-1} = \frac{adj[g_{ij}]}{\Delta},\ \Delta = \det[g_{ij}] \tag{5.2}$$

3. FIM of Equ. (5.1) is implicitly characterized by

$$(ds)^2 = \sum_{i,j=1}^{n} g_{ij}\ (d\theta^i)(d\theta^j) \tag{5.3}$$

4. The α(or $\nabla^{(\alpha)}$)-*connection* is defined for each $\alpha \epsilon \mathbb{R}$ by

$$\Gamma_{ij,k}^{(\alpha)} = \left(\frac{1-\alpha}{2}\right)(\partial_i \partial_j \partial_k(\Psi(\theta))) \tag{5.4}$$

Provided that $\Psi(\theta)$ is the potential function (c.f., definition 5.2.2) *and* $\partial_i = \frac{\partial}{\partial\theta_i}$.

Definition 5.2.4 Kullback's Divergence (KD), $K(p,q)$ (Li et al 2007)

On the manifold *M*, for any two points $p(x;\theta_p)$ and $q(x;\theta_q)$, $K(p,q)$ is given by:

$$K(p,q) = E_{\theta_p}\left[l\,n\left(\frac{p(x;\theta_p)}{q(x;\theta_q)}\right)\right] = \int p(x;\theta_p)\,l\,n\left(\frac{p(x;\theta_p)}{q(x;\theta_q)}\right)dx \qquad (5.5)$$

Provided that E_{θ_p} is value of the expectation, J-divergence, $J(p,q)$ is written as

$$J(p,q) = E_{\theta_p}\left[l\,n\left(\frac{p(x;\theta_p)}{q(x;\theta_q)}\right)\right] = \int (p(x;\theta_p) - q(x;\theta_q)\,)\; l\,n\left(\frac{p(x;\theta_p)}{q(x;\theta_q)}\right)dx \qquad (5.6)$$

Definition 5.2.5 The α **,** the $\alpha-\mathrm{Ricci}$ and the $\alpha-\mathrm{sectional\ curvature\ tensors}$ (Li et al 2007)

1. The $\alpha-\mathrm{curvature\ Riemannian}$ Tensors, $R_{ijkl}^{(\alpha)}$ for coordinates θ , $i,j,k,l,s,t =$ $1,2,3, \ldots.$ and curvature parameter, α are determined by

$$R_{ijkl}^{(\alpha)} = \left[\left(\partial_j \Gamma_{ik}^{s(\alpha)} - \partial_i \Gamma_{jk}^{s(\alpha)}\right)g_{sl} + \left(\Gamma_{jt,l}^{(\alpha)}\Gamma_{ik}^{t(\alpha)} - \Gamma_{it,l}^{(\alpha)}\Gamma_{jk}^{t(\alpha)}\right)\right],$$

with $\Gamma_{ij}^{k(\alpha)} = \Gamma_{ij,s}^{(\alpha)} g^{sk}$ (5.7)

2. The Ricci Tensors ($\alpha-\mathrm{Ricci}$ curvatures), $R_{ik}^{(\alpha)}$ are

$$R_{ik}^{(\alpha)} = R_{ijkl}^{(\alpha)} g^{jl},\ with\ R_{ijkl}^{(\alpha)}(\mathrm{c.f.}, (5.7)) \qquad (5.8)$$

3. We define $K_{ijij}^{(\alpha)}$, namely, the $\alpha-\mathrm{sectional\ curvatures}$ tensors as follows

$$K_{ijij}^{(\alpha)} = \frac{R_{ijij}^{(\alpha)}}{(g_{ii})(g_{jj}) - (g_{ij})^2}, i,j = 1,2,\ldots,n \qquad (5.9)$$

When $n = 2,$ the $\alpha-$ $\mathrm{sectional\ curvature}$ $K_{1212}^{(\alpha)}$ $= K^{(\alpha)}$is known as $\alpha-$ $\mathrm{Gaussian\ curvature}$ tensor and is given by:

$$K^{(\alpha)} = \frac{R_{1212}^{(\alpha)}}{det(g_{ij})} \qquad (5.10)$$

Definition 5.2.6 The Physical interpretation of Ricci curvature Tensor (RCT)

1. RCT could be looked at as a contraction (Loveridge 2016) of the Riemannian Tensor (c.f., (5.7)).

2. RCT (Rudelius 2012) is a descriptor of the extent of the difference between the volume of a geodesic ball on the surface and the corresponding volume of a geodesic ball in Euclidean space for an oriented Riemannian manifold M.

3. Following (Ollivier 2010), RCT provides a measure of the contraction of the evolution of volumes subjected to geodesic flow. The Bonnet Myers theorem (Ollivier 2010) states that for a positive RCT, then a sphere is less positively curved than Riemannian manifold. Furthermore, this reduces the diameter of the underlying manifold.

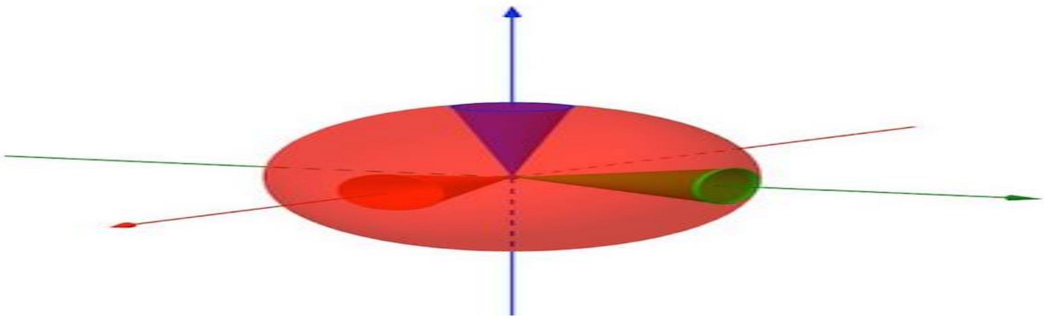


Figure 5.1: RCT offers a description of the difference between the volume of the conical regions in the manifold and their analogous conical regions in Euclidean space (Thomas 2015)

As in figure 5.1, instead of examining how the volume of a whole ball within the manifold differs from that of a ball in Euclidean space, we examine the volume of only a sliver of the ball - an angular sector or cone centred around some direction v from the ball's centre. RCT illustrates how the volume of a sliver of a ball in the manifold at point x in direction v differs from the comparable angular sector in Euclidean space.

Definition 5.2.7 The matrix exponential of Fisher information (Gunawardena 2006)

1. Given a linear system of differential equations

$$\frac{dx}{dt} = Ax \tag{5.11}$$

provided that *x* is a vector of *n*-dimensions vector and *A* is an nxn matrix. The matrix exponential, e^A given by

$$e^A = \sum_{i=0}^{\infty} \frac{A^i}{i!} = I + A + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!} + \ldots. \tag{5.12}$$

Provides the solution satisfying (5.11).

2. *A*'s characteristic polynomial is

$$\Phi(\delta) = \det(\mathrm{A} - \delta\mathrm{I}) \tag{5.13}$$

then, the set of eigen values, δ of *A* corresponding to (5.13) are:

$$\Phi(\delta) = (\delta) = \det(\mathrm{A} - \delta\mathrm{I}) = 0 \tag{5.14}$$

The eigen value δ corresponding to the eigen vector *x* satisfies:

$$Ax = \delta\, x \tag{5.15}$$

Additionally, e^A reads as

$$e^A = T\ e^D\ T^{-1} \tag{5.16}$$

D is the diagonal matrix of A's eigenvalues, and T is the matrix whose corresponding columns are A's eigenvectors.

5.3 FIM for a stable $M/G/1$ QM, and its inverse

According to (El-Affendi and Kouvatsos 1983), the maximum entropy (ME) state probability of a stable $M/G/1$ queue (c.f., Figure 5.2), under normalisation, mean queue length (MQL), L, and server utilisation, ρ(<1) constraints is determined by

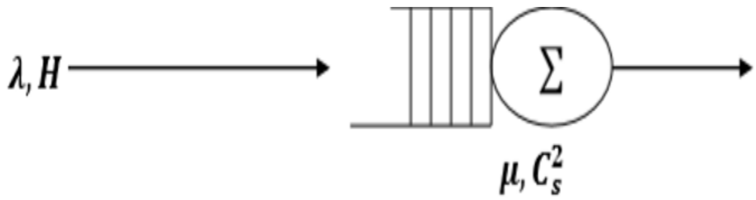


Figure 5.2: A Stable $M/G/1$ queue

$$p(n) = \begin{cases} 1-\rho, & n = 0 \\ (1-\rho)gx^n \ , & n \geq 1 \end{cases} \tag{5.17}$$

with $g = \frac{\rho^2}{(L-\rho)(1-\rho)}$, $x = \frac{L-\rho}{L}$ and $L = \frac{\rho}{2}\left(1 + \frac{1+\rho C_s^2}{1-\rho}\right)$ (MQL of Pollaczeck-Khinchin formula of a stable $M/G/1$ queue), ρ = 1 – $p(0)$ (server utilisation) and C_s^2 (SCV of the service times).

Expression (5.17) can be rewritten as

$$p(n) = \begin{cases} 1-\rho, & n = 0 \\ \frac{2\rho(\frac{1+\rho\beta}{1-\rho}-1)^{n-1}}{((\frac{1+\rho\beta}{1-\rho}+1)^n}, & n \geq 1 \end{cases} \tag{5.18}$$

Provided that $\beta = C_s^2$.

Theorem 1 For a stable $M/G/1$ QM, this is true:

(i) FIM is determined by

$$[g_{ij}] = \begin{pmatrix} \frac{1}{(1-\rho)^2} & 0 \\ 0 & \frac{-1}{(\beta+1)^2} \end{pmatrix} \tag{5.19}$$

(ii) The Fisher Information Metric is given by

$$(ds)^2 = \left(\frac{1}{(1-\rho)^2}\right)(d\rho)^2 - \frac{-1}{(\beta+1)^2}(d\beta)^2 \tag{5.20}$$

(iii) The inverse of $[g_{ij}]$ (c.f., (5.19)), $[g^{ij}]$ is devised by

$$[g^{ij}] = ([g_{ij}])^{-1} = \frac{adj[g_{ij}]}{\Delta} = \begin{pmatrix} (1-\rho)^2 & 0 \\ 0 & -(\beta+1)^2 \end{pmatrix} \quad (5.21)$$

Proof

(i) Based upon (5.17), we have two cases

Case I: For $n = 0, p(n) = 1 - \rho$. Thus,we have a one-dimensional coordinate system in this case that will be one dimensional , for which we obtain that:

$$\mathcal{L}(x;\theta) = ln\big(p(x;\theta)\big) = ln(1-\rho),\ \theta = \theta_1 = \rho \quad (5.22)$$

Thus, the potential function $\Psi(\theta)$ is:

$$\Psi(\theta) = -ln(1-\rho) \quad (5.23)$$

Hence,

$$\partial_1 = \frac{\partial\Psi}{\partial\rho} = \frac{1}{1-\rho} \quad (5.24)$$

$$\partial_1\partial_1 = \frac{\partial^2\Psi}{\partial\rho^2} = \frac{1}{(1-\rho)^2} \quad (5.25)$$

FIM is determined by:

$$[g_{ij}] = \left[\frac{\partial^2\Psi}{\partial\rho^2}\right] = \left[\frac{1}{(1-\rho)^2}\right] \quad (5.26)$$

$\left[g^{ij}\right]$ is determined by:

$$\left[g^{ij}\right] = [g_{ij}]^{-1} = [(1-\rho)^2] \quad (5.27)$$

Additionally, it holds that:

$$\Gamma_{11,1}^{(\alpha)} = (\frac{1-\alpha}{2})(\partial_1\partial_1\partial_1(\Psi(\theta))) = (\frac{1-\alpha}{(1-\rho)^3}) \quad (5.28)$$

$$\Gamma_{11}^{1(\alpha)} = \Gamma_{11,1}^{(\alpha)}(g^{11}) = \frac{1-\alpha}{(1-\rho)^3})((1-\rho)^2) = \frac{1-\alpha}{(1-\rho)} \quad (5.29)$$

$$\Gamma_{11}^{1(0)} = \frac{1}{(1-\rho)} \quad (5.30)$$

Case II: when $n \geq 1, p(n) = \frac{2\rho(\frac{1+\rho\beta}{1-\rho}-1)^{n-1}}{((\frac{1+\rho\beta}{1-\rho}+1)^{n}}$. Hence, the coordinate system in this case will be two dimensional and it clearly holds that:

$$\mathcal{L}(x;\theta) = l\,n\big(p(x;\theta)\big) = (ln2 + l\,n(\rho) + (n-1)l\,n\left(\frac{1+\rho\beta}{1-\rho} - 1\right) - nln\left(\frac{1+\rho\beta}{1-\rho} + 1\right) \quad (5.31)$$

where $\theta = (\theta_1, \theta_2) = (\rho, \beta)$.

The potential function $\Psi(\theta)$ is devised as

$$\Psi(\theta) = l\,n\left(\frac{1+\rho\beta}{1-\rho} - 1\right) - ln2 - l\,n(\rho) \quad (5.32)$$

By analogy to the above proof, after some algebraic manipulation, it clearly follows that the FIM is given by

$$[g_{ij}] = \begin{pmatrix} \frac{1}{(1-\rho)^2} & 0 \\ 0 & \frac{-1}{(\beta+1)^2} \end{pmatrix} \quad \text{(c.f., (5.19))}$$

(ii) By following analogous algebraic analysis to that of i), it can be established that the square of the arc length of ii) is given by

$$(ds)^2 = \sum_{i,j=1}^{n} g_{ij}\,(d\theta^i)(d\theta^j) = \left(\frac{1}{(1-\rho)^2}\right)(d\rho)^2 - \frac{1}{(\beta+1)^2}(d\beta)^2 \quad \text{(c.f., (5.20))}$$

(iii) Similarly, after some lengthy analytic derivations, the determinant of the FIM will be given by $\Delta = \det[g_{ij}] = -\frac{1}{(\beta+1)^2(1-\rho)^2} \neq 0$. Hence, the inverse matrix of FIM exists.

To this end, after some algebraic steps, it follows that the inverse matrix of FIM is expressed by:

$$[g^{ij}] = \begin{pmatrix} (1-\rho)^2 & 0 \\ 0 & -(\beta+1)^2 \end{pmatrix} \quad \text{(c.f., (5.21))}$$

5.4 The α(or $\nabla^{(\alpha)}$)-connection of stable $M/G/1$ QM

Following equation (5.4), it holds that:

$$\Gamma_{11,1}^{(\alpha)} = \frac{(1-\alpha)}{(1-\rho)^3} \quad (5.33)$$

By the same argument, the remaining components are obtained. By using equation (5.7), we can derive the expression:

$\Gamma_{11}^{1(\alpha)} = \Gamma_{11,1}^{(\alpha)} g^{11} + \Gamma_{11,2}^{(\alpha)} g^{21}$,

After some calculations, we have

$$\Gamma_{11}^{1(\alpha)} = \frac{1-\alpha}{(1-\rho)^3}(1-\rho)^2 = \frac{1-\alpha}{(1-\rho)}, \Gamma_{11}^{1(0)} = \frac{1}{(1-\rho)} \tag{5.34}$$

$\Gamma_{22}^{2(\alpha)} = \Gamma_{22,1}^{(\alpha)} g^{21} + \Gamma_{22,2}^{(\alpha)} g^{22}$

and after some more calculations

$$\Gamma_{22}^{2(\alpha)} = -\left(\frac{1-\alpha}{(1+\beta)^3}\right)(1+\beta)^2 = -\frac{1-\alpha}{(1+\beta)},\ \Gamma_{22}^{2(0)} = -\frac{1}{(1+\beta)} \tag{5.35}$$

Engaging the same logic, we can derive the remaining components to compute the Ricci curvature tensor, RCT of the stable $M/G/1$.

Notably, these derivations are essentially needed to get the analytic expression of RCT in section 5.6.

5.5 The KD and the JD of the stable $M/G/1$ QM

According to equation (5.5), after some few algebraic calculations, KD is expressed at $n = 0$, we have:

$$K(p,q) = E_{\theta_p}\left[ln\left(\frac{p(x;\theta_p)}{q(x;\theta_q)}\right)\right] = ln\left(\frac{1-\rho_p}{1-\rho_q}\right) \tag{5.36}$$

and for $n = 1,2,3, \ldots$.

$$K(p,q) = ln\left(\left(\frac{1-\rho_p}{1-\rho_q}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)\left(\left(\frac{\rho_q[2+\rho_q(\beta_q-1)]}{\rho_p[2+\rho_p(\beta_p-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_p} \tag{5.37}$$

where L_p is the corresponding MQL of Pollaczeck-Khinchin Formula of a stable $M/G/1$ QM at $p, i.e.\ L_p = \sum_{n=0}^{\infty} np(n)$ (c.f., (5.17)).

This could be summarized in the more compact form:

$$K(p,q) = \begin{cases} ln\left(\frac{1-\rho_p}{1-\rho_q}\right), & n = 0 \\ ln\left(\left(\frac{1-\rho_p}{1-\rho_q}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)\left(\left(\frac{\rho_q[2+\rho_q(\beta_q-1)]}{\rho_p[2+\rho_p(\beta_p-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_p}, & n = 1,2,3,.. \end{cases} \quad (5.38)$$

More interestingly, in a similar fashion, the JD is derived in the two possible cases :

Case 1: $n = 0$

$$J(p,q) = K(p,q) + K(q,p) = ln\left(\frac{1-\rho_p}{1-\rho_q}\right) + ln\left(\frac{1-\rho_q}{1-\rho_p}\right) = 0 \quad (5.39)$$

Case 2: $n = 1,2,3,...$

$J(p,q) = K(p,q) + K(q,p) =$

$$= \left(ln\left(\left(\frac{1-\rho_p}{1-\rho_q}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)\left(\left(\frac{\rho_q[2+\rho_q(\beta_q-1)]}{\rho_p[2+\rho_p(\beta_p-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_p} \right.$$

$$\left. + \ ln\left(\left(\frac{1-\rho_q}{1-\rho_p}\right)\left(\frac{1+\beta_p}{1+\beta_q}\right)\right)\left(\left(\frac{\rho_p[2+\rho_p(\beta_p-1)]}{\rho_q[2+\rho_q(\beta_q-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_q} \right)$$

$$= \left(\left(\left(\frac{\rho_q[2+\rho_q(\beta_q-1)]}{\rho_p[2+\rho_p(\beta_p-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_p} + \left(\left(\frac{\rho_p[2+\rho_p(\beta_p-1)]}{\rho_q[2+\rho_q(\beta_q-1)]}\right)\left(\frac{1+\beta_q}{1+\beta_p}\right)\right)^{L_q}\right) \neq 0 \quad (5.40)$$

Equation (5.40) shows that $J(p,q)$ of stable $M/G/1$ QM is generally non-zero. This is never the case at the initial steady state phase, when $n = 0$.

5.6 The α –Gaussian and RICCI curvature tensors of the stable $M/G/1$ QM

In this section, it is devised that the stable $M/G/1$ QM has a zero α –Gaussian curvature tensor, $K^{(\alpha)}$ for all the values of the curvature parameter α as well as having a non-zero Ricci curvature tensor.

Theorem 2 The stable $M/G/1$ QM has

i) a zero α –Gaussian curvature tensor and ii) Has a non-zero Ricci curvature tensor

Proof

Case i), by definition (5.10), it is enough to show that the α −Gaussian curvature:

$$K^{(\alpha)} = \frac{R_{1212}^{(\alpha)}}{det(g_{ij})} = 0 \tag{5.41}$$

To this end, the expression $R_{ijkl}^{(\alpha)}$ is given by:

$$R_{ijkl}^{(\alpha)} = \left[\left(\partial_j \Gamma_{ik}^{s(\alpha)} - \partial_i \Gamma_{jk}^{s(\alpha)}\right) g_{sl} + \left(\Gamma_{jt,l}^{(\alpha)} \Gamma_{ik}^{t(\alpha)} - \Gamma_{it,l}^{(\alpha)} \Gamma_{jk}^{t(\alpha)}\right)\right],$$

and α is the curvature parameter with $\Gamma_{ij}^{k(\alpha)} = \Gamma_{ij,s}^{(\alpha)} g^{sk}$(c.f., (5.7)

Specifically,

$$R_{1212}^{(\alpha)} = \left[\left(\partial_2\left(\Gamma_{11}^{1(\alpha)} + \Gamma_{11}^{2(\alpha)}\right) - \partial_1(\Gamma_{21}^{1(\alpha)} + \Gamma_{21}^{2(\alpha)}\right))(g_{12} + g_{22})\right.$$

$$\left. + \left(\Gamma_{21,2}^{(\alpha)} \Gamma_{11}^{1(\alpha)} + \Gamma_{22,2}^{(\alpha)} \Gamma_{11}^{2(\alpha)}) - (\Gamma_{11,2}^{(\alpha)} \Gamma_{21}^{1(\alpha)} + \Gamma_{12,2}^{(\alpha)} \Gamma_{21}^{2(\alpha)}\right)\right]$$

$$= \left[\left(\frac{\partial}{\partial\beta}\left(\frac{1-\alpha}{1-\rho} + 0\right) - \frac{\partial}{\partial\rho}\left(0 - \frac{1}{(\beta+1)^2}\right))\right))(g_{12} + g_{22}) + 0\right] = 0 \tag{5.42}$$

$det(g_{ij}) = -\frac{1}{(\beta+1)^2(1-\rho)^2} \neq 0$. Hence, $K^{(\alpha)} = \frac{R_{1212}^{(\alpha)}}{det(g_{ij})} = 0$. Thus i) follows.

Case ii) We need to show that at least one of the components of $\alpha - \mathrm{RCs}$, $R_{ik}^{(\alpha)}$ are given by:

$$R_{ik}^{(\alpha)} = R_{ijkl}^{(\alpha)} g^{jl}, i, j, k, l = 1,2,3, \ldots., n \qquad \text{(c.f.,(5.9))}$$

is non-zero. To this end, $R_{11}^{(\alpha)}$ is expressed by:

$$R_{11}^{(\alpha)} = R_{1212}^{(\alpha)} g^{11} + R_{1112}^{(\alpha)} g^{12} + R_{1211}^{(\alpha)} g^{21} + R_{1212}^{(\alpha)} g^{22} \tag{5.43}$$

Engaging the same procedure as in (5.43), after some manipulation, one gets

$$R_{11}^{(\alpha)} = 0 \tag{5.44}$$

Similarly,

$$R_{12}^{(\alpha)} = R_{22}^{(\alpha)} = 0 \tag{5.45}$$

Additionally, $R_{21}^{(\alpha)}$ is determined by:

$$R_{21}^{(\alpha)} = R_{2111}^{(\alpha)} g^{11} + R_{2112}^{(\alpha)} g^{12} + R_{2211}^{(\alpha)} g^{21} + R_{2212}^{(\alpha)} g^{22} \tag{5.46}$$

Moreover, $g^{11} = (1-\rho)^2$, $g^{22} = -(\beta+1)^2$

Following some lengthy calculations, it is obtained that:

$$R_{21}^{(\alpha)} = -\frac{(1-\alpha)^2}{(1-\rho)^4}(1-\rho)^2 + 0 + 0 - (\beta+1)^2(0) = -\frac{(1-\alpha)^2}{(1-\rho)^2} \neq 0 \tag{5.47}$$

Hence, ii) follows.

From (5.47), it holds that:

$$R_{21}^{(0)} = -\frac{1}{(1-\rho)^2} \tag{5.48}$$

Clearly, $R_{21}^{(0)}$ is ρ –dependent. As $\rho \to 1$, $R_{21}^{(0)} \to -\infty$. This illustrates the influential impact of (stability)(instability) of the two dimensional stable $M/G/1$ QM. Clearly, figure 5.3 reveals that the stability (equivalently, $1 > \rho > 0$) of the underlying stable $M/G/1$ QM impacts RCT to be an increasing function in $(\text{server utilization})\rho$. More interestingly, figure 5.4 strongly supports the fact that the instability phase(equivalently, $\rho \geq 1$) of the stable $M/G/1$ QM enforces RCT to be a decreasing function in ρ.

More fundamentally, this illustrates how RCT impacts the stability dynamics of $M/G/1$ QM.

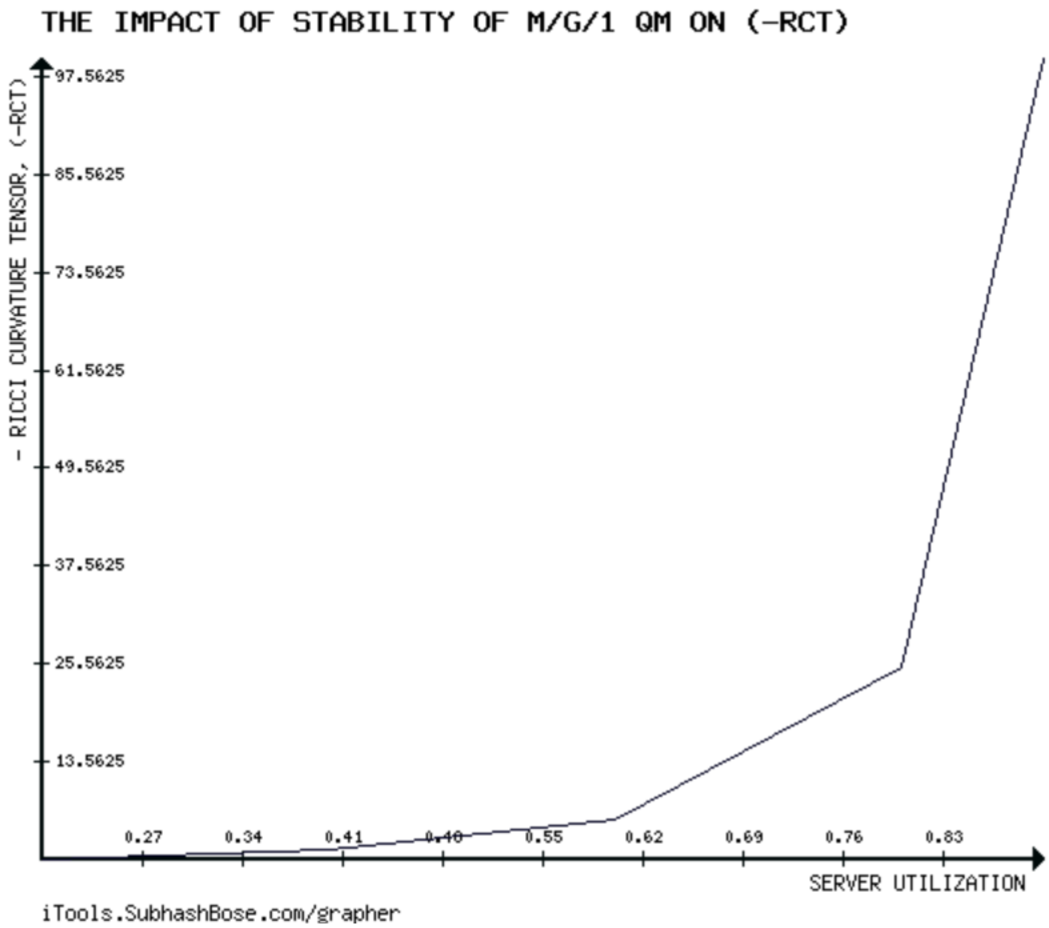


Figure 5.3: The $M/G/1$ QM's stability phase forces Ricci Curvature (RCT) to be an increasing function in ρ

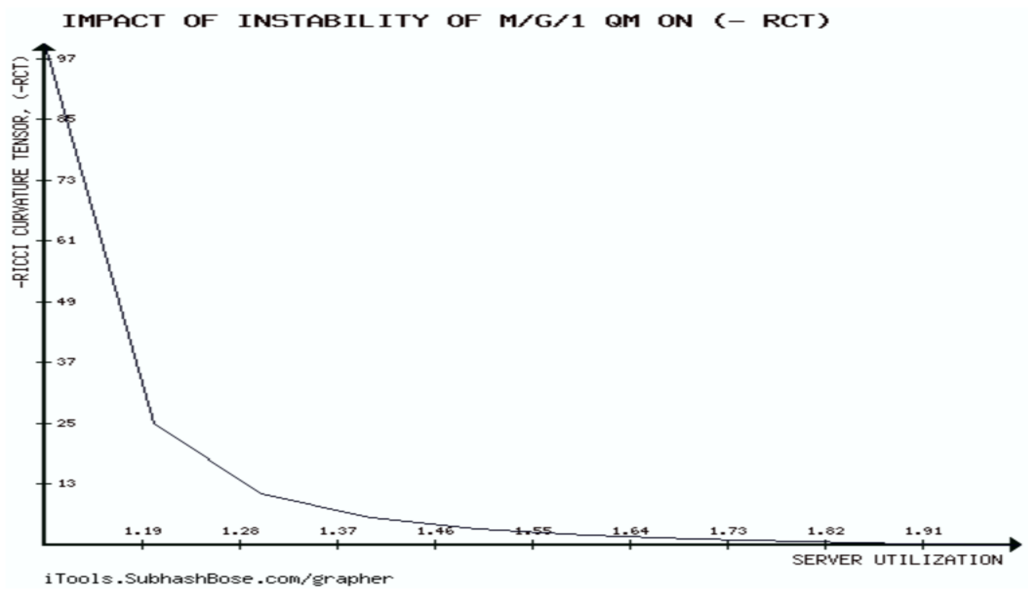


Figure 5.4: The $M/G/1$ QM's instability phase forces Ricci Curvature (RCT) to be a decreasing function in ρ.

The presented data portrait in both figures 5.3 and 5.4 is collectively combined in figure 5.5.

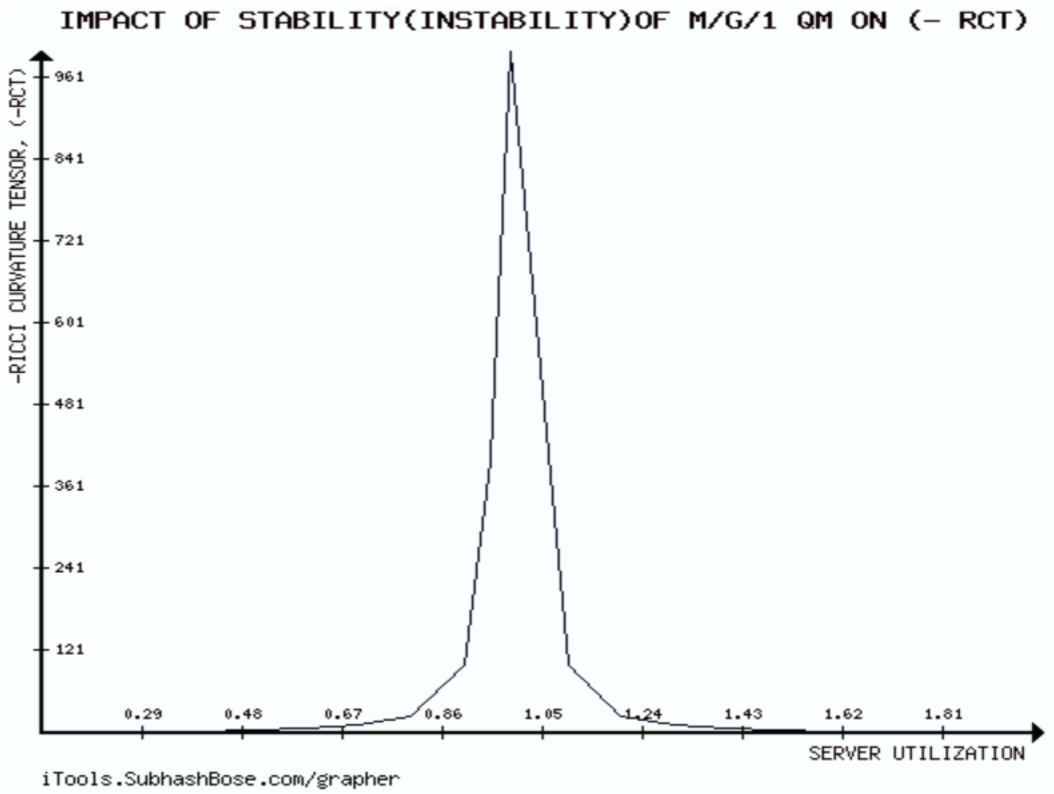


Figure 5.5 : The influential impact of (stability)instability phase of the $M/G/1$ QM on the behaviour of Ricci Curvature (RCT).

In the following theorem, $\boldsymbol{e}^{\boldsymbol{FIM}_{\boldsymbol{stable}\ \boldsymbol{M/G/1}\ \boldsymbol{QM}}}$ serves as the exponential matrix of the FIM of stable $M/G/1$ QM.

5.7 $\boldsymbol{e}^{\boldsymbol{FIM}_{\boldsymbol{stable}\ \boldsymbol{M/G/1}\ \boldsymbol{QM}}}$

Theorem 3 $\boldsymbol{e}^{\boldsymbol{FIM}_{\boldsymbol{stable}\ \boldsymbol{M/G/1}\ \boldsymbol{QM}}}$ solves a differential equation of the form:

$$\frac{dx}{dt} = \mathrm{A}x \tag{5.49}$$

Proof

By equation (5.19) of Theorem 1, we have

$$[g_{ij}]=\begin{pmatrix} \frac{1}{(1-\rho)^2} & 0 \\ 0 & \frac{-1}{(\beta+1)^2} \end{pmatrix} \qquad \text{(c.f., (5.19))}$$

Re-writing $[g_{ij}]$, yields

$$[g_{ij}] = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, a = \left(\frac{1}{(1-\rho)^2}\right), b = \frac{-1}{(\beta+1)^2} \tag{5.50}$$

Hence

$$\Phi(\delta) = (\delta) = \det\left([g_{ij}] - \delta\mathrm{I}\right) = \det \begin{pmatrix} a-\delta & 0 \\ 0 & b-\delta \end{pmatrix} = 0 \tag{5.51}$$

Thus, we have

$$\delta^2 - (a+b)\delta + ab = 0 \tag{5.52}$$

Hence, the eigenvalues are given by $\delta_{1,2} = a, b$. Therefore, the diagonal matrix *D* is given by

$$D = \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix} \tag{5.53}$$

For $\delta_{1,2} = a, b$, the corresponding eigen vectors are $\binom{1}{0}$, $\binom{0}{1}$. Clearly, it follows that the matrix:

$$T = T^{-1} = \text{unity matrix } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.54}$$

Hence, $\boldsymbol{e^{FIM_{stable\ M/G/1\ QM}}}$ is devised by:

$$e^A = Te^DT^{-1} = \begin{pmatrix} e^a & 0 \\ 0 & e^b \end{pmatrix} \tag{5.55}$$

By (5.5), it follows that:

$$\frac{dx}{dt} = Ax \qquad \text{(c.f., (5.49))}$$

5.8 Chapter Summary

Information geometry has been successfully used to obtain the orthogonal decomposition of probability distributions with exponential or mixed groups have a natural hierarchical structure(see Amari, 2001). Decomposing a stochastic dependency onto several random variables is a common example. They have a complex network of dependencies in general. A broad class hierarchy specifies orthogonal decomposition, which includes both exponential and mixed families. As an example, we decompose higher-order Markov chain dependencies into the sum of the dependencies of various lower-order Markov chains. A single server, such as the $M/G\ /1$ system, is straightforward and can serve as a preliminary model (Hamasha et al 2016).

Based on the preceding discussion, our new approach to revealing the significant impact of IG on queuing theory clarified the lost connection. The stability of the queuing theory (Rachev 1989) problem is concerned with the continuity of the mapping F from set U of the input flow to set V of the output flow. First, if U and V have a metric structure, we use probability theory to estimate the coefficient of F- continuity. Following that, we evaluate the error term in the approximation of the input flow with a simpler one while observing some functionals of the empirical input flow distribution. This demonstrates the power of a new approach in determining the exact stability and instability phases of the underlying $M/G/1$ queueing system for the first time.

Classic Queueing theory is revolutionized by our innovative techniques since we are considering queues as manifolds, considering α as a curvature parameter and an underlying stable $M/G/1$ QM connection parameter. In the metric connection (see Jefferson 2018), the inner product is preserved by translation of the two vectors, thus preserving its importance in Riemannian geometry. Both metric and symmetric connections are Riemannian manifolds and generally exist infinitely.

More surprisingly, in general, the non-metricity of statistical manifolds' natural metrics is justified by the Riemannian connection on a curved surface that is defined by the special case α = 0, $\nabla^{(\alpha)}$is from the Fisher metric point of view (however, $\nabla^{(\alpha)}$)can be any value of α. On the other hand, it is symmetric.

The exponentiation of a matrix (Lee 1950)corresponds to their respective Jordan block's powers. It is a fact that this interpretation applies to e^X and all analytic functions f that that are applicable to matrices. It is also beneficial to rethink a matrix exponential as a "system solution for ordinary differential equations (ODEs)".

Potentially as a future work, we can develop further advancements to info- geometric queuing theory by analysing the stability of $G/G/1$ queue manifold (Dodson, 2005; Kouvatsos1988). Additionally, we can investigate applications of manifolds and information geometry to various statistical manifolds and transient queueing systems.

6. The Upper and Lower Bounds of The Data Information Length of Transient $M/M/\infty$ Queuing System

In this chapter, an exposition of a novel link between Information Length Theory (ILT) and Transient Queueing Systems (TQSs) is undertaken by deriving both upper and lower bounds of the data information length of a transient $M/M/\infty$ queuing system. In this context, it is revealed that if $\rho(t)$ serves as the time-dependent server utilization of the transient $M/M/\infty$ queuing system, then the latter obtained upper and lower bounds($UB(n,t), LB(n,t))$ respectively) are both $(n, \rho(t))$-dependent, $n = 0,1,2, \dots$.. Additionally, a typical numerical experiment is conjectured to illustrate the significant impact of time on behaviour of the devised $UB(n,t)$ and $LB(n,t)$ for different values of n.

6.1 The Poisson process

One of the most popular counting (Pishro-Nik 2014)methods is the Poisson process. It is typically employed in situations when we are counting the occurrences of specific events that seem to occur at a certain rate but are completely random (without a certain structure). For instance, let's say we know from past data that there are two earthquakes that happen in a specific region every month. The timing of earthquakes appears to be completely random except for this information. Thus, we draw the conclusion that the Poisson process may serve as a useful earthquake model. Models have

- Photons arriving on a photodiode.

- The quantity of auto accidents at a location or in a region.
- The position of users in a wireless network.
- Requests for certain publications on a web server.
- The start of conflicts.

A random process is called Poisson process (Pishro-Nik 2014)with the rate μ if it satisfies the following definition:

For a fixed $\mu > 0$,$t \in (0, \infty)$, we define a counting process to be Poissonian with rates μ when it satisfies the following:

1. **$N(0) = 0$;**
2. **the underlying increments of $N(t)$ are independent,**
3. **within any interval having a length $\vartheta > 0$, the associated number of arrivals must follow a Poissonian $(\mu\vartheta)$ distribution.**

Specifically, if

$$T_n = \sum_{i=1}^{n} X_i,\ T_n \sim Gamma(n, \mu) \tag{6.1}$$

provided that $X_i's$ serve as to be randomized independent variables satisfying $Exponential(\mu)$ variables. Then,

$$E[T_n] = \frac{1}{\mu} \tag{6.2}$$

$$Var[T_n] = \frac{1}{\mu^2} \tag{6.3}$$

This provides a simulation approach for a Poissonian process of a rate μ. We start with the generated $X_i \sim Exponential(\mu)$ to obtain the corresponding service times

$$T_1 = X_1, \tag{6.4}$$

$$T_2 = X_1 + X_2, \tag{6.5}$$

$$T_3 = X_1 + X_2 + X_3, \tag{6.6}$$

.

.

.

Additionally, T_n must follow (6.2) and (6.3).

6.2. The transient $M/M/\infty$ queue and IL

The $M/M/\infty$ queue (c., f., Harrison et al., 1992) is a multi-server queueing model used in queueing theory, an emerging applied probabilistic discipline, where every arrival receives instant service and does not wait, as demonstrated by figure 6.1. According to (Kumar et al. 2014; Kulkarni 2016), the Poisson distributed service time with mean μ and mean arrival rate i*s* λ in the $M/M/\infty$ queueing system's transient probability is given by:

$$p_n(t) = \frac{[\frac{\lambda}{\mu}(1-e^{-\mu t})]^n}{n!} \exp\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\}, \; n = 0,1,2, \ldots. \tag{6.7}$$

Notably, as $t \to \infty$, $p_n(t)$ of Equ. (6.7) converges to

$$p_n = \frac{\rho^n}{n!} e^{-\rho}, \; n = 0,1,2, \ldots. \tag{6.8}$$

ρ serves as the server utilization of the underlying queue.

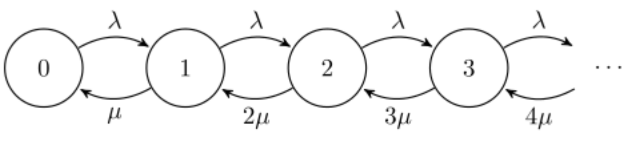


Figure 6.1. The state space diagram for the $M/M/\infty$ chain

Recalling that the information length (IL) is defined to be:

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)} = \int_0^t \sqrt{(\varepsilon(t_1))}dt_1 \quad \text{(c.f., 2.34))}$$

$$\varepsilon(t_1) = \int_{\mathbb{R}^n} \left(\frac{1}{p(x,t_1)}[\frac{\partial p(x,t_1)}{\partial t_1}]^2\right)dx \quad \text{(c.f., 2.35))}$$

If $\mathcal{L}_{M/M/\infty}$ serves as the IL of the transient $M/M/\infty$ queuing system, then the following theorems are devoted to the derivation of the lower and upper bounds of $\mathcal{L}_{M/M/\infty}$,namely ($LB(n,t), UB(n,t)$respectively).

Theorem 6.1

The IL of the transient $M/M/\infty$ queuing system **,** $\mathcal{L}_{M/M/\infty}$ satisfies the following inequality:

$$\frac{2(n+1)(\rho(t))^{\frac{3}{2}}}{3\sqrt{n!}} > \mathcal{L}_{M/M/\infty}(t) > \left(\frac{\sigma\mu^n}{2^n n!}\right)\int_0^t t^n e^{-\rho(t)}\rho^{\cdot}(t)(\rho(t))^n dt \quad (6.9)$$

Proof

We have

$$p_n(x,t) = \frac{[\frac{\lambda}{\mu}(1-e^{-\mu t})]^n}{n!}\exp\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\} \quad \text{(c.f., (6.7))}$$

$$= \frac{x^n}{n!}e^{-x}, x(t) = \frac{\lambda}{\mu}(1-e^{-\mu t}) = \rho(t)(1-e^{-\mu t}) \ \left(since\ \rho(t) = \frac{\lambda(t)}{\mu}\right) \quad (6.10)$$

By the definition, we have $\mu > 0,$ which implies that $e^{-\mu t} < 1.$ Hence, $x(t) < \rho(t)$ follows. Then, we have the real number σ ∈(0,1) satisfying:

$$x(t) = \sigma\rho(t) \quad (6.11)$$

Consequently,

$$\frac{\partial p_n(t)}{\partial t} = \frac{\sigma x^{n-1}\rho^{\cdot}(t)}{n!}e^{-x}(n-x) \quad (6.12)$$

$$\frac{\left[\frac{\partial p_n(x,t)}{\partial t}\right]^2}{p_n(x,t)} = \frac{\left(\frac{\sigma x^{n-1}\rho^{\cdot}(t)}{n!}e^{-x}(n-x)\right)^2}{\frac{x^n}{n!}e^{-x}}$$

$$= \frac{\sigma^2 x^{n-2}\rho^{\cdot 2}(t)}{n!}e^{-x}(n-x)^2$$

$$= \frac{\sigma^2 x^{n-2}\rho^{\cdot 2}(t)}{n!}e^{-x}(n^2-2nx+x^2)$$

$$\leq (\frac{\sigma^2\rho^{\cdot 2}(t)}{n!})(n^2+1) \text{ (Since } e^{-x} \leq 1, x \in (0,1))$$

$$< (\frac{\rho^{\cdot 2}(t)}{n!})(n^2+1)(\text{since } \sigma \in (0,1)) \tag{6.13}$$

Hence, it follows that:

$$\mathcal{L}_{M/M/\infty} < \int_0^t \sqrt{(}\int \frac{(n+1)^2\rho^{\cdot 2}(t)}{n!}dx)\, dt$$

$$= \frac{(n+1)}{\sqrt{n!}}\int_0^t (\sqrt{\int}\, dx])\ \rho^{\cdot}(t)dt$$

$$= \frac{(n+1)}{\sqrt{n!}}\int_0^t (\sqrt{x}))\rho^{\cdot}(t)dt$$

$$= \frac{(n+1)}{\sqrt{n!}}\int_0^t \sqrt{\frac{\lambda}{\xi}(1-e^{-\mu t})}\rho^{\cdot}(t)dt$$

$$< \frac{(n+1)}{\sqrt{n!}}\int_0^t \sqrt{(}\rho(t)\rho^{\cdot}(t)dt \text{ (Since } e^{-\mu t} \in (0,1)$$

$$= \frac{2(n+1)(\rho(t))^{\frac{3}{2}}}{3\sqrt{n!}} \tag{6.14}$$

On the other hand, by $x < \rho(t) and\ x < 1$, we have

$$\frac{\left[\frac{\partial p_n(x,t)}{\partial t}\right]^2}{p_n(x,t)} = \frac{\sigma^2 x^{n-2}\rho^{\cdot 2}(t)}{n!}e^{-x}(n-x)^2 > e^{-2x}\left(\frac{\sigma^2 x^{n-2}\rho^{\cdot 2}(t)}{n!}\right)(n-1)^2$$

$$> e^{-2\rho(t)}\left(\frac{\sigma^2 x^{2n-1}\rho^{\cdot 2}(t)}{n!}\right)(n-1)^2$$

$$> \sigma^2 e^{-2\rho(t)}\rho^{\cdot 2}(t)\left(\frac{x^{2n-1}}{(n!)^2}\right)(n-1)^2 \qquad (6.15)$$

Thus, it follows that:

$$\left(\sqrt{\int \frac{\left[\frac{\partial p_n(x,t)}{\partial t}\right]^2}{p_n(x,t)}dx}\right) > \left[\int \sigma^2 e^{-2\rho(t)}\rho^{\cdot 2}(t)\left(\frac{x^{2n-1}}{(n!)^2}\right)(n-1)^2\right]dx > \sigma\rho^{\cdot}(t)\, e^{-\rho(t)}\left(\frac{x^n}{n!}\right) \qquad (6.16)$$

Hence,

$$\mathcal{L}_{M/M/\infty}(t) > \left(\frac{\sigma}{n!}\right)\int_0^t \rho^{\cdot}(t)e^{-\rho(t)}\left(\rho(t)(1-e^{-\mu t})\right)^n dt$$

$$> \left(\frac{\sigma}{n!}\right)\int_0^t e^{-\rho(t)} \quad \text{(Since, } (1-e^{-\mu t}) > \frac{\mu t}{2} \qquad \text{(c. f., Ovler, et al. 2010))}$$

$$= \left(\frac{\sigma\mu^n}{2^n n!}\right)\int_0^t t^n e^{-\rho(t)}\rho^{\cdot}(t)(\rho(t))^n dt \qquad (6.17)$$

Therefore,

$$\frac{2(n+1)(\rho(t))^{\frac{3}{2}}}{3\sqrt{n!}} > \mathcal{L}_{M/M/\infty}(t) > \left(\frac{\sigma\mu^n}{2^n n!}\right)\int_0^t t^n e^{-\rho(t)}\rho^{\cdot}(t)(\rho(t))^n dt \qquad \text{(c.f., (6.9))}$$

In what follows, an illustration of the temporal impact as well as the potential impact of n on the behaviour of both $LB(n,t)$and $UB(n,t)$,

$$UB(n,t) = \frac{2(n+1)(\rho(t))^{\frac{3}{2}}}{3\sqrt{n!}} \text{ and } LB(n,t) = \left(\frac{\sigma\mu^n}{2^n n!}\right)\int_0^t t^n e^{-\rho(t)}\rho^{\cdot}(t)(\rho(t))^n dt \qquad (6.18)$$

6.3 Numerical Experiments

By choosing $\sigma = \frac{1}{2}, \rho(t) = \frac{1}{t}, \mu$ =2, it can be easily verified that these proposed choices,$UB(n,t)$ and $LB(n,t)$ will read:

$$UB(n,t) = \frac{2(n+1)}{3(t)^{\frac{3}{2}}\sqrt{(n!)}} \text{ and } LB(n,t) = \left(-\frac{1}{2(n!)}\right)\int_0^t e^{-\frac{1}{t}}\frac{1}{t^2}dt = \left(-\frac{e^{-\frac{1}{t}}}{2(n!)}\right) \quad (6.19)$$

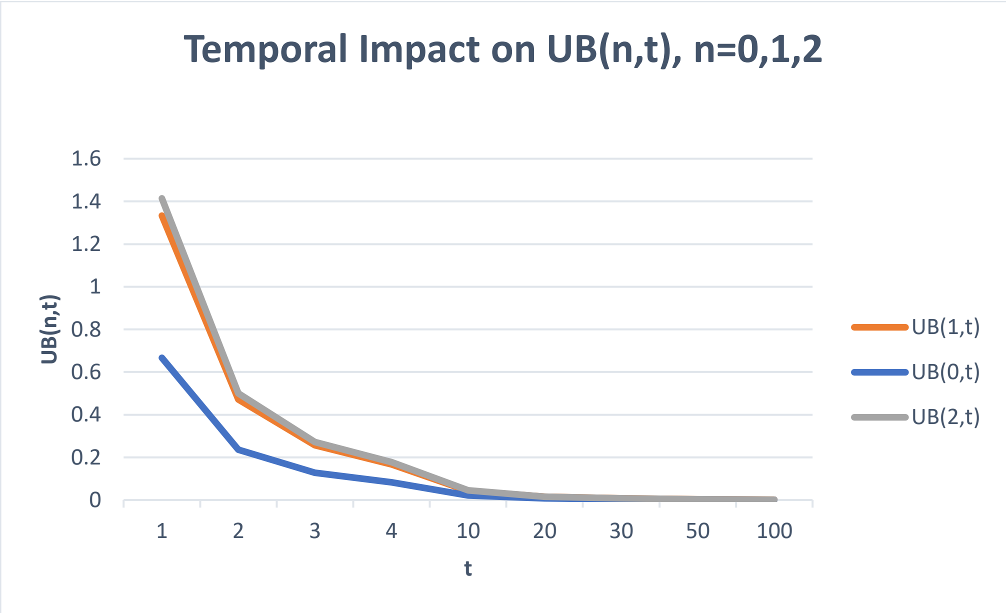


Figure 6.2. Significant Impact of time and n on $UB(n,t)$

It is clear from Figure 6.2, that because of the progressive increase of time, $UB(n,t), n = 0,1,2$ are decreasing functions in time. Moreover, for each recorded value of time, by increasing the value of n, $UB(n,t), n = 0,1,2$ are increasing functions. More interestingly, this reveals that $UB(n,t)$ acts as decreasing function in time and an increasing function with respect to n. Fundamentally, the increase of time and $n$impacts $UB(n,t)$ to depict longer heavy tails. This translates to seeing the longest heavy tails for $UB(2,t)$, and these tails become shorter for $UB(1,t)$ and the shortest would be for $UB(0,t)$.

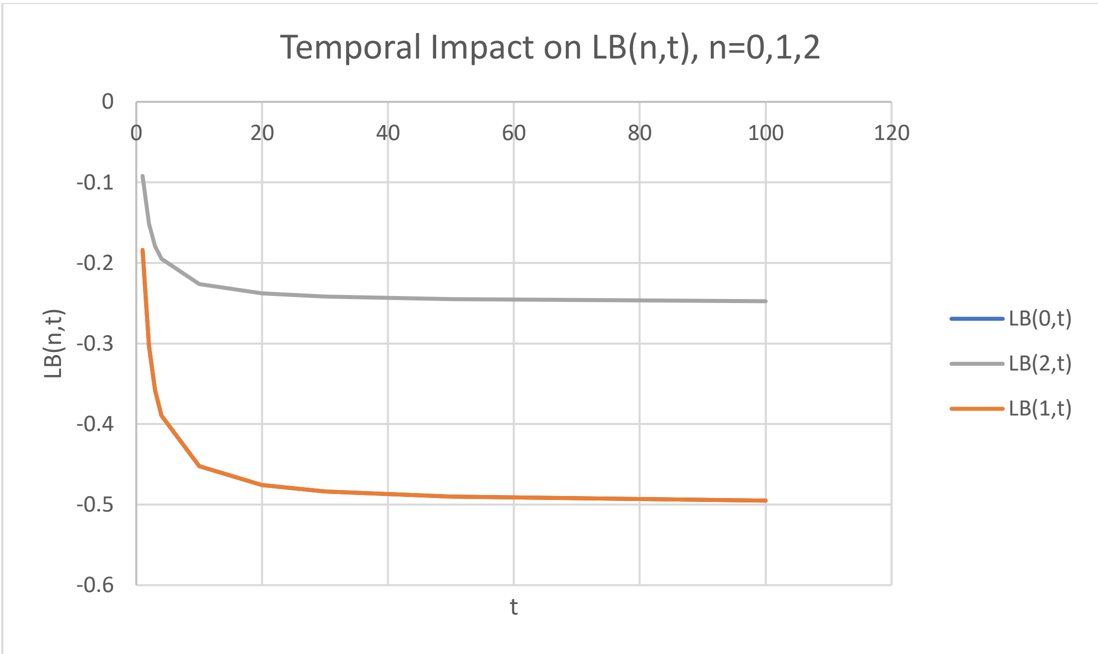


Figure 6.3. Significant Impact of time and n on $LB(n,t)$

Reading Figure 6.3, $LB(n,t)$is decreasing in time and increasing as n increases. More potentially, the graph representation of $LB(n,t)$ produces shorter heavy tails by the increase of time and n. This shows a complete converse scenario in comparison to the recorded heavy tails phenomena in $UB(n,t)$.

In mathematical terms, as time becomes sufficiently large($t \to \infty$), it follows that

$$lim_{t\to\infty} UB(n,t) = lim_{t\to\infty} \frac{2(n+1)}{3(t)^{\frac{3}{2}}\sqrt{n!}} = 0 = lim_{t\to\infty} \left(-\frac{e^{-\frac{1}{t}}}{2(n!)}\right) = \; lim_{t\to\infty} \left(-\frac{1}{2(n!)}\right) \int_0^t e^{-\frac{1}{t}} \frac{1}{t^2} dt$$

$$= \; lim_{t\to\infty} LB(n,t) \tag{6.20}$$

Engaging the findings of Eqs. (6.9) and (6.20), it holds that as time reaches infinity, $n = 0{,}1{,}2$

$$lim_{t\to\infty}\mathcal{L}_{M/M/\infty}\ (t) = 0 \qquad (6.21)$$

Notably , it is shown that for $M/M/\infty$ transient queueing system(Kumar et al. 2014; Kulkarni 2016) as $t \to \infty$, the correspondent steady state probability density function, p_n is devised by

$$p_n = \frac{\rho^n}{n!}e^{-\rho},\ n = 0,1,2, \dots. \qquad \text{(c.f., (6.8))}$$

Since p_n(c.f., (6.8)) does not depend on time, then we have by the definition of IL, (c.f., (2.34), (2.35)), that the corresponding stability phase of $M/M/\infty$ queueing system has an underlying zero information length, that is:

$$lim_{t\to\infty}\mathcal{L}_{M/M/\infty}\ (t) = 0, n = 0,1,2,\ M/M/\infty \text{ is stable} \qquad (6.22)$$

Clearly, equations (6.21) and (6.22) are equivalent. This provides a strong validation of both obtained mathematical and numerical results. The strategy of the proof can be extended to the remaining values of $n = 3,4,5, \dots..$,which will show that $UB(n,t)$ will start to decrease in both (n,t), i.e, the behavioural trend will reverse in terms of the Increasability phase in n. More interestingly, it can be verified that $LB(n,t)$ will never change its behavioural trend in (n,t) by being temporarily decreasing and increasing with the respective increase of n.

In a more detailed account, communicating (Guichard 2017), it could be analytically demonstrated that $UB(n,t)$ and $LB(n,t)$ (c.f., 6.19) are both (increasing in t)(decreasing in n) by showing that :

I) $\frac{\partial UB(n,t)}{\partial t} < 0, \frac{\partial UB(n,t)}{\partial n} > 0$ (6.23)

II) $\frac{\partial LB(n,t)}{\partial t} < 0, \frac{\partial LB(n,t)}{\partial n} > 0$ (6.24)

We have

$$\frac{\partial UB(n,t)}{\partial t} = \frac{\partial}{\partial t}\left(\frac{2(n+1)}{3(t)^{\frac{3}{2}}\sqrt{(n!)}}\right) = -\frac{(n+1)}{(t)^{\frac{5}{2}}\sqrt{n!}} < 0 \text{ for all } t > 0 \tag{6.25}$$

Additionally, engaging (Peng 2020), the Stirling formula to compute $n!$ is written as

$$n! \sim \sqrt{(2\pi n)}\left(\frac{n}{e}\right)^{n}, \qquad n = 3,4,\dots \tag{6.26}$$

$$\frac{\partial UB(n,t)}{\partial n} \sim \frac{2}{3(t)^{\frac{3}{2}}(2\pi)^{\frac{1}{4}}}\frac{\partial}{\partial n}\left(\frac{(n+1)}{\left(\frac{n}{e}\right)^{\frac{n}{2}}n^{\frac{1}{2}}}\right) = \frac{2}{3(t)^{\frac{3}{2}}(2\pi)^{\frac{1}{4}}}\frac{\partial}{\partial n}\left((n^{\frac{1}{2}} + n^{-\frac{1}{2}})\left(\frac{n}{e}\right)^{-\frac{n}{2}}\right)$$

$$= \frac{2}{3(t)^{\frac{3}{2}}(2\pi)^{\frac{1}{4}}}\left(\left(\frac{1}{2}n^{-\frac{1}{2}} - \frac{1}{2}n^{-\frac{3}{2}}\right)\left(\frac{n}{e}\right)^{-\frac{n}{2}} + (n^{\frac{1}{2}} + n^{-\frac{1}{2}})\left(-\frac{1}{2}(lnn - 1) - \frac{1}{2}\right)\left(\left(\frac{n}{e}\right)^{-\frac{n}{2}}\right)\right)$$

$$= \frac{\left(\frac{n}{e}\right)^{-\frac{n}{2}}}{3(t)^{\frac{3}{2}}(2\pi)^{\frac{1}{4}}}\left(\left(n^{-\frac{1}{2}}(1 - lnn) - n^{-\frac{3}{2}}\right) - (n^{\frac{1}{2}})lnn\right) < 0, n = 3,4,\dots \tag{6.27}$$

This proves (I).

On the other hand, we have

$$\frac{\partial LB(n,t)}{\partial t} = \frac{\partial}{\partial t}\left(-\frac{e^{-\frac{1}{t}}}{2(n!)}\right) = -\left(\frac{e^{-\frac{1}{t}}}{2(n!)t^2}\right) < 0 \text{ for all } t > 0, n = 3,4,\dots \tag{6.28}$$

Additionally,

$$\frac{\partial LB(n,t)}{\partial n} \sim \left(-\frac{e^{-\frac{1}{t}}}{2\sqrt{(2\pi)}}\right)\frac{\partial}{\partial n}\left(n^{-\frac{1}{2}}\left(\frac{n}{e}\right)^{-n}\right)$$

$$= \left(-\frac{e^{-\frac{1}{t}}\left(\frac{n}{e}\right)^{-\frac{n}{2}}}{2\sqrt{(2\pi)}}\right)\left(-\frac{1}{2}n^{-\frac{3}{2}} - n^{-\frac{1}{2}}lnn\right)$$

$$= \left(\frac{e^{-\frac{1}{t}} n^{-\frac{3}{2}} \left(\frac{n}{e}\right)^{-\frac{n}{2}}}{4\sqrt{(2\pi)}} \right) (1 + 2nlnn) > 0 \text{ for all } t > 0, n = 3,4, \dots \qquad (6.29)$$

Hence, (II) follows.

Fundamentally, we have demonstrated with strong supporting mathematical evidence that the undertaken experiments agree with the analytic proofs.

6.4 Chapter Summary

This chapter contributes to the establishment of Information Length Theory of Transient Queueing Systems. The novel mathematical derivations are undertaken by finding the integral formula of the information length of the transient $M/M/\infty$ queuing system. Because of the complexity to derive the closed form result of the later integral formula, both the upper and the lower bounds of that integral, namely $UB(n,t), LB(n,t))$ were derived.

More interestingly, it is observed that $UB(n,t), LB(n,t))$ are both $(n, \rho(t))$-dependent, $n = 0,1,2, \dots \dots$, with $\rho(t)$ to define the time-dependent server utilization of the transient $M/M/\infty$ queuing system. Moreover, these analytic findings were validated numerically.

7.Conclusions

The properties of the family of Rényi's generalised entropies (RGEs) which are classified as being information theoretic, notably as probabilistic techniques for inductive inference, have primarily been documented in the continuous-time domain in the past state of available knowledge. An original extension of these properties into the discrete-time domain, a generalisation of an analytical result on the interpretation of maximum entropy (ME) formalism as a consistency requirement, and a determination of the Probability Vector Updates (PV-updates), based on a reworking of prior information-theoretic results on minimum cross entropy, are devised in this chapter.

Reflecting on the key aims of this thesis, in chapter four, new addition to knowledge in the field of stable queues, specifically focusing on the information theoretic impact of a non-extensive parameter has been revealed. This impact has led to the development of two new state probabilities, the Rényian and Tsallisian closed form expressions, for the underlying stable $M/G/1$ queueing system. The derived solutions have been proven to be exact and credible, satisfying three consistency axioms while defying one due to the non-extensivity impact.

Chapter five highlights a significant advancement in queueing theory by integrating it with various mathematical fields such as differential geometry, information theory, matrix theory, and information geometry (IG). This integration allows for the application of info-geometric techniques, which provide a powerful approach to studying queue stability and revolutionizing classical queueing theory through innovative IG methods. This interdisciplinary approach opens new possibilities for understanding and optimizing queueing systems.

Chapter six discusses the connection between Information Length Theory (ILT) and Transient Queueing Systems (TQSs). It explains how upper and lower bounds of the data information length can be derived for a transient $M/M/\infty$ queuing system, based on the time-dependent server utilization $\rho(t)$. The impact of time on the behavior of the derived bounds is illustrated through a numerical experiment, and potential real-life applications of the transient $M/M/\infty$ queuing system are mentioned.

7.1. Limitations

The fundamental drawback of this study endeavour in the latter situation is the time independence anticipated in stable queueing systems. Potentially, it is not possible to compute the corresponding information length(IL) for stable queueing systems based on time independence, perhaps this could lead to the re-definition of a novel form of information length formalism that is not based on time but is based on the underlying queueing parametrization.

Generalized $\mathbb{Z}$ entropy function has been used to characterise non-extensive entropy maximisation axiomatically (Mageed and Zhang 2022(b)). The new ME solutions developed in this thesis are special cases of the heavy-tailed solutions derived using non-extensive entropies, subject to the same prior information constraints, because both Rényian and Tsallisian entropy functionals are special cases of non-extensive Generalized $\mathbb{Z}$ entropy (Mageed and Zhang 2022(b)). The research presented in this thesis thus provides examples that may be utilised to test and create new heavy-tailed ME queueing systems.

7.2. Future directions

Future goals for this research include the provision of info-geometric queueing theory (IGQT), information length theory of transient queueing systems, and unique methodologies developed in this thesis as steppingstones toward the modern unification of information theory with queuing theory (ILTTQSs).

The following are some challenging open problems that have arisen because of this research:

1.It is to be recalled experimentally, for the undertaken extended properties for the Generalized Rényian entropy functional in the discrete time domain, H_q^L, it has been proven that for $q \in (-1, \infty)$ $\lim_{q \to a} H_q^L = H_a^L$, namely, the limit theory.The challenging unsolved open problem till current is to prove or disprove that

$$\lim_{q \to a} H_q^L = H_a^L \text{ for } q \in (-\infty, 1)$$

2. More complicated open questions appear in the horizon such as the provision of these extended properties in discrete time domains for higher order generalized entropies that generalize Rényian entropy functional such as Generalized $\mathbb{Z}$ entropy and others in the literature, as well as proving or disproving the limit theory for the Generalized $\mathbb{Z}$ entropy corresponding to any real value for the information-theoretic parameter q (Mageed and Zhang 2022(b)).

3.Is it feasible to obtain EME and NME formalisms (consequently, short-range, and long-range interactions, respectively) for a more complicated philosophical forms of entropic measures that generalize this current work (Mageed and Zhang 2022(b)).

4.If adding higher level moments to the prescribed set of constraints , what form will the new derivations take and what type of GE will be generated. If this is the case, can we generate both service time PDF and CDF that make our solution exact? It is expected that these newly derived service time PDF and CDF will involve parameters that correspond to the added higher moments within the prescribed constraints.

5. Following the foundational achievements of this thesis, there are several next phase advancements of IG analysis of all the available categories of queueing systems in the literature. More interestingly, the provision of info-geometric analysis of the dynamics of transient queueing systems.

6. Following the novel derivations of the information length approach for the transient $M/M/\infty$ queuing system, another development to the theory will be the unification of IL theory with several phenomena in physics, for example, the IL interpretation of photon statistics and generalized Brownian motion. Additionally, undertaking further advancements to IL for other transient queueing systems.

Appendix A

Detailed proofs of chapter three

A.1. The proof of the uniqueness property (3.4.1)

Herewith, there are only three possibilities:

First possibility: $q > 0$

We know that $(\sum_{i=1}^{J=2^n} x_i^{q+1})^{-1/q}$ is maximal if and only if $\sum_{i=1}^{J=2^n} x_i^{q+1}$ is minimal.

Firstly, the convexity of x^{q+1} is required to be demonstrated on the unit interval [0,1]. Hence, $d^2g/dx^2 = q(q+1)x^{q-1}$ and this implies $d^2g/dx^2 > 0$ for all $x \in$(0,1]. Secondly, it is a must to prove the convexity of $\sum_{i=1}^{J=2^n} x_i^{q+1}$ is on $\mathbb{Q}^L$, i.e. that for all $0 \leq \lambda \leq 1$, any two elements $a^{\rightarrow}, b^{\rightarrow}$ in $\mathbb{Q}^L$, $\sum_i(\lambda a_i + (1-\lambda)b_i)^{q+1} \leq \lambda \sum_i a_i{}^{q+1} + (1-\lambda)\sum_i b_i{}^{q+1}$, this reads as

$$\sum_i(\lambda a_i + (1-\lambda)b_i)^{q+1} - \left(\lambda \sum_i a_i{}^{q+1} + (1-\lambda)\sum_i b_i{}^{q+1}\right) \leq 0 \quad \text{for all } a^{\rightarrow}, b^{\rightarrow} \in \mathbb{Q}^L \tag{A.1}$$

(=0 if and only if $a_i = b_i$).Engaging the convexity of x^{q+1}, one gets

$$\lambda a^{q+1} + (1-\lambda)b^{q+1} \geq (\lambda a + (1-\lambda)b)^{q+1} \tag{A.2}$$

Equality holds in (A.2) whenever $a = b$. Hence, we get :

$$\sum_i(\lambda a_i + (1-\lambda)b_i)^{q+1} \leq \left(\lambda \sum_i a_i{}^{q+1} + (1-\lambda)\sum_i b_i{}^{q+1}\right) \tag{A.3}$$

as required because of the continuity of $\sum_i x_i{}^{q+1}$ is over $\mathbb{Q}^L$ together with the compactness of $V^L(K)$, it follows that $a^{\rightarrow} \epsilon\, V^L(K)$ is a minmum point. Let a distinct point $b^{\rightarrow} \epsilon\, V^L(K)$ be another assumed minimum, this implies $\sum_i a_i{}^{q+1} = \sum_i b_i{}^{q+1}$. Because of the strict convexity of $\sum_i x_i{}^{q+1}$,it could be obtained that $\sum_i(\frac{a_i+b_i}{2})^{q+1} < \frac{1}{2}\sum_i a_i{}^{q+1} + \frac{1}{2}\sum_i b_i{}^{q+1} = \sum_i a_i{}^{q+1}$(contradiction). Thus, there exists a unique maximal point of $(\sum_i x_i{}^{q+1})^{-1/q}$ whenever $q > 0$.

Second possibility: Following the definition, H_0^L = ME holds.

Third possibility: $0 > q > -1$,the proof is like case 1.

A.2. The proof of the null limit (3.4.2)

It is needed to demonstrate that : (i) if given $\varepsilon > 0\ \exists\delta > 0$ such that if $|q| < \delta, x^{\to}\epsilon\mathbb{Q}^L,\ |log(\sum_i x_i{}^{q+1})^{-\frac{1}{q}} + \sum_{i=1}^J x_i log x_i\,| < \varepsilon$ which is the same as proving $|(1/q)log\sum_i x_i{}^{q+1} - \sum_{i=1}^J x_i log x_i\,| < \varepsilon$. (ii) $a^{\to(q)} = < a_1^{(q)}, \dots, a_J^{(q)} >$ is the point in $\mathrm{V^L(K)}$ at which $\log(\sum_\mathrm{i}\mathrm{x_i}{}^{\mathrm{q}+1})^{-\frac{1}{\mathrm{q}}}$ is maximum, then $lim_{q\to 0}\ a^{\to(q)} = a^{\to(0)}$. To show (i), we know that by using the Taylor expansion for $log(\sum_i x_i{}^{q+1})^{-\frac{1}{q}}$ around $q = 0, log\ (\sum_{i=1}^J x_i{}^{q+1}) - log\ (\sum_{i=1}^J x_i) = (q\left[\frac{d}{dt}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=0} +$ $\frac{q^2}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}})$

where $0 < \theta_{x^{\to}} < q$. We know that *log* $\sum_i x_i = 0$, since $\sum_i x_i = 1$ and consequently, one gets:

$$log\left(\sum_{i=1}^J x_i{}^{q+1}\right) = (q\left[\frac{d}{dt}\left(log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=0} + \frac{q^2}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}})$$

$$\left(q\left[\frac{d}{dt}\left(log\left(\sum_{i=1}^J x_i{}^{r+1}\right)\right)\right]_{t=0} + \frac{q^2}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}}\right) = (q\left[\left(\sum_i x_i{}^{t+1}\, log\, x_i/\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=0} + \frac{q^2}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}}) = (q\left[\left(\sum_i x_i\, log\, x_i/\left(\sum_{i=1}^J x_i\right)\right)\right]_{t=0} + \frac{q^2}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}})$$

which implies

$$\frac{log\left(\sum_{i=1}^J x_i{}^{q+1}\right)}{q} - \sum_i x_i\, log x_i = \frac{q}{2}\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^J x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\to}}} \tag{A.4}$$

We are to show that : $\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\rightarrow}}}$ has an independent of $x^{\rightarrow}$ upper bound. As already known,

$$\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right) = \frac{\left(\sum_i x_i{}^{t+1}(logx_i)^2\right)-\left(\sum_i x_i{}^{t+1}logx_i\right)^2}{\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)^2}$$

This implies:

$$\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\rightarrow}}} \leq \left[\frac{\left(\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1}(logx_i)^2\right)}{\sum_{i=1}^{J} x_i{}^{\theta_{x^{\rightarrow}}+1}} + \frac{\left(\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1}logx_i\right)^2}{\left(\sum_{i=1}^{J} x_i{}^{\theta_{x^{\rightarrow}}+1}\right)^2}\right] \qquad \text{(A.5)}$$

From elementary calculus we have

$$lim_{x\searrow 0}xlogx = lim_{x\searrow 0}x(logx)^2 \qquad \text{(A.6)}$$

This will imply that:

$$f(x) = xlogx, g(x) = x(logx)^2, for\ all\ x\ such\ that\ 0 < x \leq 1, f(0) = g(0) = 0$$

$$\left|\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1}(logx_i)^2\right| \leq \sum_i \left|x_i{}^{\theta_{x^{\rightarrow}}+1}(logx_i)^2\right| = \sum_i \left|x_i{}^{\theta_{x^{\rightarrow}}}\right| x_i(logx_i)^2\Big| \leq \sum_i x_i(logx_i)^2$$

The continuity requirement holds for both f and g. Additionally, we have

(We can drop $\theta_{x^{\rightarrow}}$ because $|x_i{}^{\theta_{x^{\rightarrow}}}| \leq 1$ since $0 \leq x \leq 1, 0 \leq \theta_{x^{\rightarrow}} < r$)

$$\leq \sum_{i=1}^{2^n} M_2 = 2^n M_2 \qquad \text{(A.7)}$$

Similarly, we have

$$\left(\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1}logx_i\right)^2 \leq \left(\sum_i x_i logx_i\right)^2 \leq \left(\sum_{i=1}^{2^n} M_1\right)^2 = 2^{2n}M_1^2 \qquad \text{(A.8)}$$

By $\sum_{i=1}^{2^n} x_i = 1$, then $x_i \geq \frac{1}{2^n}$for some x_i . equating the remaining $x_i'\,s$ to zero, yields

$$\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1} \geq \left(\frac{1}{2^n}\right)^{\theta_{x^{\rightarrow}}+1} \qquad \text{(A.9)}$$

By (A.7) and (A.9), one gets:

$$[\frac{\left(\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1}(logx_i)^2\right)}{\sum_{i=1}^{J} x_i{}^{\theta_{x^{\rightarrow}}+1}}] \leq 2^n M_2/\left(\frac{1}{2^n}\right)^{\theta_{x^{\rightarrow}}+1} = (2^n)^{\theta_{x^{\rightarrow}}+2}\, M_2 \qquad \text{(A.10)}$$

Also, we have by (A.8) and (A.9) that:

$$|\frac{(\sum_i x_i{}^{\theta_{x^{\rightarrow}}+1} logx_i)^2}{\left(\sum_{i=1}^{J} x_i{}^{\theta_{x^{\rightarrow}}+1}\right)^2}| \leq 2^{2n} M_1^2/\left(\frac{1}{2^n}\right)^{\theta_{x^{\rightarrow}}+1} = (2^n)^{\theta_{x^{\rightarrow}}+3}\, M_1^2 \qquad \text{(A.11)}$$

Thus, we have by (A.10) and (A.11) that:

$$\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\rightarrow}}} \leq [(2^n)^{\theta_{x^{\rightarrow}}+2}\, M_2 + (2^n)^{\theta_{x^{\rightarrow}}+3}\, M_1^2]$$

$$< (2^n)^{\theta_{x^{\rightarrow}}+3}(M_2 + M_1^2) \qquad \text{(A.12)}$$

By $q \searrow 0$, let $r \leq 2$. Then, $\theta_{x^{\rightarrow}} < 2$. Replacing $\theta_{x^{\rightarrow}} + 3$ within (A.12), it follows that $\left[\frac{d^2}{dt^2}\left(\log\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right)\right]_{t=\theta_{x^{\rightarrow}}} < 2^{5n}(M_2 + M_1^2\ )$ (A.13). Choosing M to be such that

$$M = (M_2 + M_1^2)$$

We have from (A.4) and (A.13) that $\left|\frac{log\left(\sum_{i=1}^{J} x_i{}^{q+1}\right)}{q} - \sum_i x_i\, logx_i\right| < \left(\frac{q}{2}\right) M < (\delta/2)M.$

Now choosing our δ to be such that $\delta \leq 2\varepsilon/M$ implies the needed outcome.

Following the same logic, we can derive the proof when $q < 0$. As for (ii), Assuming the contrary, then there exists a subsequence $a^{\rightarrow(q_n)}$ when $q_n \searrow 0$ for which $lim_{n\to\infty}\, a^{\rightarrow(q_n)} = b^{\rightarrow} \neq a^{\rightarrow(0)}$. Hence, ∃ a number $\eta \epsilon (0,1)$ such that $\sum_i b_i\, logb_i - \sum_i a_i^{(0)}\, loga_i^{(0)} > \eta > 0$ and by the compactness of $V^L(K)$, we have $b^{\rightarrow} \in V^L(K)$. If we choose $\delta > 0$such that $|q| < \delta$, we have:

$$\left|\frac{log\left(\sum_{i=1}^{J} x_i{}^{q+1}\right)}{q} - \sum_i x_i\, logx_i\right| < \frac{\eta}{13}$$

for all $x^{\rightarrow} \in \mathbb{D}^L$. Now, since $q_n \searrow 0$, pick q_n such that $0 < |q_n| < \delta$,

$$\left|\sum_i a_i^{(q_n)} \, loga_i^{(q_n)} - \sum_i b_i \, logb_i\right| < \frac{\eta}{13}$$

From this we obtain

$$\frac{1}{q_n}\log\left(\sum_i (a_i^{(0)})^{1+q_n}\right) \le \sum_i (a_i^{(0)}) \, log(a_i^{(0)}) + \frac{\eta}{13}$$

$$< \sum_i b_i \, logb_i - \frac{12\eta}{13}$$

$$< \left(\sum_i (a_i^{(q_n)}) \, log(a_i^{(q_n)})\right) + \frac{\eta}{13} - \frac{12\eta}{13}$$

$$\le \left(\frac{1}{r_n}\log\left(\sum_i (a_i^{(0)})^{1+q_n}\right)\right) + \frac{\eta}{13} - \frac{12\eta}{13}$$

$$< \left(\frac{1}{q_n}\log\left(\sum_i (a_i^{(0)})^{1+q_n}\right)\right)$$

Which results in the necessary contradiction. For $q_n \nearrow 0$, a largely equivalent proof holds true.

A.3. The full proof of the limit theorem (3.4.3)

(i) The uniform convergence can be demonstrated if we proved that if given $\varepsilon > 0$ $\exists\, \delta > 0$ such that :

$|q - a| < \delta$, then $|\left(\sum_{i=1}^{J} x_i{}^{q+1}\right) - \left(\sum_{i=1}^{J} x_i{}^{a+1}\right)| < \varepsilon$ for all $x^{\rightarrow} \in \mathbb{Q}^L$.

(ii)If $b^{\rightarrow(q)} = < b_1^{(q)}, \ldots, b_J^{(q)} >$ is the point in $V^L(K)$at which $\sum_{i=1}^{J} x_i{}^{q+1}$ is maximal (minimal), then:

$lim_{q \rightarrow a} \; b^{\rightarrow(q)} = \; b^{\rightarrow(a)}$. To show (i), we know that by using the Taylor expansion for $\sum_{i=1}^{J} x_i{}^{q+1}$ around q =*a* one gets

$\left(\sum_{\mathrm{i}=1}^{\mathrm{J}} \mathrm{x_i}^{\mathrm{q}+1}\right)$-$\left(\sum_{\mathrm{i}=1}^{\mathrm{J}} \mathrm{x_i}^{\mathrm{a}+1}\right)$

$$= (q-a)\left[\frac{d}{dt}\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right]_{t=a} + \frac{(q-a)^2}{2}\left[\frac{d^2}{dt^2}\left(\sum_{i=1}^{J} x_i{}^{t+1}\right)\right]_{t=\theta_{x^{\rightarrow}}} \qquad \text{(A.13)}$$

where $0 < \theta_{x^{\rightarrow}} < q$. Thus we have

$$\left(\sum_{\mathrm{i}=1}^{\mathrm{J}} \mathrm{x_i}^{\mathrm{q}+1}\right)\text{-}\left(\sum_{\mathrm{i}=1}^{\mathrm{J}} \mathrm{x_i}^{\mathrm{a}+1}\right) = (q-a)\left[\left(\sum_{i=1}^{J} x_i{}^{a+1}\, log x_i\right)\right]_{t=a} +$$

$$\frac{(q-a)^2}{2}\left[\left(\sum_{i=1}^{J} x_i{}^{t+1}(log x_i)^2\right)\right]_{t=\theta_{x^{\rightarrow}}} \quad \text{(A.14)}$$

We are to show that $\left(\sum_{i=1}^{J} x_i{}^{a+1}\, log x_i\right)$ and

$\left[\left(\sum_{i=1}^{J} x_i{}^{t+1}(log x_i)^2\right)\right]_{t=\theta_{x^{\rightarrow}}}$ have an independent of $x^{\rightarrow}$ upper bound. As known that that for any real number $\beta > 0$ it holds that:

$$lim_{x\searrow 0} x^{\beta}(log x)^2 \qquad \text{(A.15)}$$

This will imply that:

$f(x) = x^{a+1}(log x)^2$ and $g(x) = x^{a+1} log x$ for $0 < x \le 1$, $f(0) = 0 = g(0) =$

are continuous functions and hence are bounded on [0,1] by M_3 and M_4 respectively. Hence, we obtain:

$$|\left(\sum_{i=1}^{J} x_i{}^{\theta_{x^{\rightarrow}}+1}(log x_i)^2\right)| \le \sum_i |x_i{}^{\theta_{x^{\rightarrow}}+1}(log x_i)^2| = \sum_i |x_i{}^{\theta_{x^{\rightarrow}}-a} x_i{}^{a+1}(log x_i)^2| =$$

$$\sum_i |x_i{}^{\theta_{x^{\rightarrow}}-a}||x_i{}^{a+1}(log x_i)^2|$$

(We can drop $\theta_{x^{\rightarrow}} - a$ here because $|x_i{}^{\theta_{x^{\rightarrow}}-a}| \le 1$ since $0 \le \mathrm{x} \le 1$, $0 < \theta_{\mathrm{x}^{\rightarrow}} < \mathrm{q}$)

$$\le \sum_i |x_i{}^{a+1}(log x_i)^2| = 2^n M_3 \qquad \text{(A.16)}$$

Similarly, we have

$\sum_i |{x_i}^{a+1}\, log x_i\,| \leq 2^n M_4$ (A.17)

Therefore, we obtain:

$$\left|\left(\sum_{i=1}^{J} {x_i}^{q+1}\right) - \left(\sum_{i=1}^{J} {x_i}^{a+1}\right)\right| \leq \left[|q-a|2^n M_4 + \left(\frac{|q-a|^2}{2}\right)2^n M_3\right]$$

$$< \left[2^n \delta M_4 + \left(\frac{\delta^2}{2}\right)2^n M_3\right]$$

$$< 2^n(\delta M_4 + \delta M_3) = \ 2^n\delta(M_3 + M_4) \quad \text{(A.18)}$$

Now choosing our δ to be such that $\delta \leq \varepsilon/M$, the proof follows. Engaging the same argument, the proof holds in the case $\mathrm{q} < \mathrm{a}$. To prove (ii), assuming the contrary.Then ∃ a subsequence $b^{\rightarrow(r_n)}$ when $q_n \searrow a$ such that $lim_{n\to\infty} b^{\rightarrow(q_n)} = \ c^{\rightarrow} \neq b^{\rightarrow(a)}$, ∃ a positive number $\eta\epsilon\ (0{,}1)$ such that we have

$$\sum\nolimits_i c_i^{a+1} - \sum\nolimits_i (b_i^{(a)})^{a+1} > \eta > 0$$

and by the compactness of $V^L(K)$, $c^{\rightarrow}\epsilon\ \ V^L(K)$(for $-1 < a < 0$ it is enough to reverse the above inequality. Choosing $\delta > 0$ such that for |q-a| < δ we have $\left|\left(\sum_{i=1}^{J} {x_i}^{q+1}\right) - \left(\sum_{i=1}^{J} {x_i}^{a+1}\right)\right| < \frac{\eta}{13}$. Now since $\mathrm{q_n} \searrow \mathrm{a}$, pick $\mathrm{q_n}$ such that $|q_n - a| < \delta$ and from this we obtain:

$$\sum_i \left(b_i^{(q_n)}\right)^{1+q_n} > \sum_i \left(b_i^{(q_n)}\right)^{1+r_n} + \frac{\eta}{13} + \frac{\eta}{13} - \left(\frac{12\eta}{13}\right)$$

$$\geq \sum_i \left(b_i^{(q_n)}\right)^{a+1} + \frac{\eta}{13} - \left(\frac{12\eta}{13}\right)$$

$$\geq \sum_i c_i^{a+1} - \left(\frac{12\eta}{13}\right)$$

$$\geq \frac{\eta}{13} > \sum_i c_i^{a+1} - \sum_i \left(b_i^{(a)}\right)^{a+1} \quad \text{(Contradiction)}$$

A similar proof works for $\mathrm{q_n} \nearrow \mathrm{a}$.

A.4. The full proof of the property of principle of irrelevant information

We know that H_0^L = ME satisfies irrelevant information by part 2 of theorem (3.4.2). Let $q > 0$ and let K_1 be the set of constraints $x + y + z = 1(so\ Bel(\neg p_1 \wedge \neg p_2) = 0\ automatically) Ny + z = f$,where $N > 1$ is large and $(1 + \left(\frac{N}{N-1}\right)^{\frac{1}{q}})f = 1$ consequently,

$$d/dy[(1 - f + (N-1)y)^{q+1}] + y^{q+1} + (f - Ny)^{q+1}]_{y=0} = 0$$

From which it follows that

$$H_q^L(K_1)\left(p_1 \bigwedge \neg p_2\right) = 0 \qquad (A.19)$$

Now, let $x_1 = Bel(p_1 \wedge p_2 \wedge p_3), y_1 = Bel(p_1 \wedge \neg p_2 \wedge p_3), z_1 = Bel(\neg p_1 \wedge \neg p_2 \wedge p_3), x_2 = Bel(p_1 \wedge p_2 \wedge \neg p_3), y_2 = Bel(p_1 \wedge \neg p_2 \wedge \neg p_3), z_2 = Bel(\neg p_1 \wedge \neg p_2 \wedge \neg p_3)$, and let K_2 be the set of constraints $K_1 + Bel(p_3) = d$, i.e.,that $x_1 + x_2 + y_1 + y_2 + z_1 + z_2 = 1, N(y_1 + y_2) + z_1 + z_2 = f, x_1 + y_1 + z_1 = d$,where$1 > d > f$ satisfies,

$-f^q + (d - f)^q - (1 - d)^q = 0^*$[* notice that there is such a *d* by the intermediate value theorem since, replaced by 1, $-f^q + (1 - f)^q - (1 - 1)^q > 0(notice\ f < 1/2)$] whilst with *d* replaced by $f, -f^q + (f - f)^q - (1 - f)^q < 0.$] Now suppose H_q^L satisfied irrelevant information. Then, by language invariance, $H_q^L(K_2)(p_1 \wedge \neg p_2) = H_q^L(K_1)(p_1 \wedge \neg p_2) = 0$ by (A.19). Therefore, by obstinacy $H_q^L(K_2) = H_q^L(K_3)$, where K_3 is the set of constraints $y_1 + y_2 = 0$ implying that $x_1 = d - f + z_2$, $z_1 + z_2 = f$, $x_1 + z_1 = d\ giving\ x_2 = 1 - d - z_2$.Now,

$$\frac{d}{dz_2}[(f - z_2)^{q+1} + {z_2}^{q+1} + (d - f + z_2)^{q+1} + (1 - d - z_2)^{q+1}]_{z_2=0}$$

$$= (r + 1)[-f^q + (d - f)^q - (1 - d)^q] = 0$$

$H_q^L(K_2)(\mathrm{p}_1 \wedge \neg\mathrm{p}_2 \wedge \neg\mathrm{p}_3) = 0. \mathrm{So}, H_q^L(K_2) = H_q^L(K_3)$ is the determined solution to K_3 using

$$x_1 = d - f, x_2 = 1 - d, y_1 = y_2 = 0, \ \ z_1 = f, z_2 = 0 \tag{A.20}$$

However, $\frac{d}{dt}[(d - f + (N-1)t)^{q+1} + (1-t)^{q+1} + t^{q+1} + (f - Nt)^{q+1}]_{t=0}$

$$= (q+1)[(N-1)(d-f)^q - Nf^q] = 0 \ < (q+1)[(N-1)(1-f)^q - Nf^q] = 0$$

From which it follows that if $\varepsilon > 0$ is very small and $x_1' = d - f + (N-1)\varepsilon$, $x_2' = 1 - d, y_1' = \varepsilon, y_2' = 0, z_1' = f - N\varepsilon, z_2' = 0$,then

$$[(x_1')^{q+1} + (x_2')^{q+1} + (y_1')^{q+1} + (y_2')^{q+1} + (z_1')^{q+1} + (z_2')^{q+1}]$$

$$< (d-f)^q + (1-f)^q + 0^{q+1} + 0^{q+1} + f^q + 0^{q+1}$$

But this contradicts (A.20) since $x_1', x_2', y_1', y_2', z_1', z_2' = 0$ is a solution of K_2. It follows that H_q^L cannot satisfy irrelevant information. The following example shows that $H_{-1/2}^L$ fails to satisfy irrelevant information:- Suppose that K_1 is $x + y + z = 1, 3y + z = 1$, with $x = Bel(\mathrm{p}_1 \wedge \mathrm{p}_2), y = Bel(\mathrm{p}_1 \wedge \neg\mathrm{p}_2), z = Bel(\neg\mathrm{p}_1 \wedge \neg\mathrm{p}_2)$. Then $H_{-1/2}^L(K_1)$ is the solution to these equations given by $z = 1 - 3y, x = 2y, \ and \ y = \frac{5+3\sqrt{2}}{42}$. Now let K_2 be $K_1 + Bel(\mathrm{p}_3)= \frac{5+3\sqrt{2}}{42}$, so the equivalence between $x_1 + x_2 + y_1 + y_2 + z_1 + z_2 = 1$, $3(y_1 + y_2) + z_1 + z_2 = f, \ z_1 = \frac{5+3\sqrt{2}}{42}$ and K_2 holds ,where $x_1 = Bel(\mathrm{p}_1 \wedge \mathrm{p}_2 \wedge \mathrm{p}_3)$ etc. Taking y_1, y_2, z_1 as the independent variables then $H_{-1/2}^L(K_1)$ will be the solution of

$$\frac{\partial D}{\partial y_1} = \frac{\partial D}{\partial y_2} = \frac{\partial D}{\partial z_1} = 0 \tag{A.21}$$

since $H_{-1/2}^L$ satisfies open-mindedness, where:

$$D = \sqrt{x_1} + \sqrt{x_2} + \sqrt{y_1} + \sqrt{y_2} + \sqrt{z_1} + \sqrt{z_2}$$

Now assume that $H_{-1/2}^L$ did satisfy irrelevant information. Then $H_{-1/2}^L(K_2)(\mathrm{p}_1 \wedge \neg\mathrm{p}_2) = y_1 + y_2 = \frac{5+3\sqrt{2}}{42}$.

So by obstinacy $H^L_{-1/2}(K_2)$ = $H^L_{-1/2}(K_3)$ where $K_3 = K_2 + Bel(\mathrm{p}_1 \wedge \neg\mathrm{p}_2)$= $\frac{5+3\sqrt{2}}{42}$, in other words $y_1 = \frac{5+3\sqrt{2}}{42} - y_2$ will hold for $H^L_{-1/2}(K_2)$ in particular the equations in (A.21) viz:-

$$\frac{-1}{\sqrt{(a-y_1-z_1)}} + \frac{3}{\sqrt{(3y_1+2y_2-a+z_1)}} + \frac{1}{\sqrt{y_1}} - \frac{3}{\sqrt{(1-3y_1-3y_2-z_1)}} = 0,$$

$$\frac{2}{\sqrt{(3y_1+2y_2-a+z_1)}} - \frac{3}{\sqrt{(1-3(y_1+y_2)-z_1)}} + \frac{1}{\sqrt{y_1}} = 0,$$

$\frac{-1}{\sqrt{(a-y_1-z_1)}} + \frac{3}{\sqrt{(3y_1+2y_2-a+z_1)}} + \frac{1}{\sqrt{z_1}} - \frac{3}{\sqrt{(1-3y_1-3y_2-z_1)}} = 0$, where $= \frac{5+3\sqrt{2}}{42}$, become with the substitution

$y_1 = a - y_2$,

$$\frac{-1}{\sqrt{(y_2-z_1)}} + \frac{3}{\sqrt{(2a-y_2+z_1)}} + \frac{1}{\sqrt{(a-y_2)}} = 0 \tag{A.22}$$

$$\frac{2}{\sqrt{(2a-y_2+z_1)}} + \frac{3}{\sqrt{(y_1)}} - \frac{3}{\sqrt{(1-3a-z_1)}} = 0 \tag{A.23}$$

$$\frac{2}{\sqrt{(y_2-z_1)}} + \frac{3}{\sqrt{(z_1)}} - \frac{2}{\sqrt{(1-3a-z_1)}} = 0 \tag{A.24}$$

Now (A.22) + (A.24) have the unique solution for $0 \leq y_2, \quad z_1 \leq 1$ of $y_2 = 0.16275594, z_1 = 0.07534572 (to\ at\ least\ 5dp)$

However, substituting these values in the left-hand side of (A.23) gives value 0.01272 instead of zero, giving the required contradiction! From these results we would conjecture that irrelevant information fails for H_q whenever $-1 < q < 0$.

A.5. The full proof of relativisation principle

Suppose that K_1 , $K_2 \in CL$, $0 < c < 1$ and

$K_1 = \{Bel(\emptyset) = c\} + \left\{\sum_{j=1}^{r} a_{ji}\, Bel\left(\frac{\theta_i}{\emptyset}\right) = b_i \,\middle|\, i = 1, \dots, m\right\}, K_2 = K_1 + \left\{\sum_{j=1}^{q} e_{ji}\, Bel\left(\frac{\psi_i}{\neg\emptyset}\right) = f_i \,\middle|\, i = 1, \dots, s\right\}.$

Also, suppose that $K_2' = K_1' = \{Bel(\emptyset) = 1\} + \left\{\sum_{j=1}^{r} a_{ji}\, Bel(\theta_i \wedge \phi) = b_i \;\middle|\; i = 1, \dots, m\right\}$,

$K_1'' = \{Bel(\neg\emptyset) = 1\}$,$K''_2 = \{Bel(\neg\emptyset) = 1\} + \left\{\sum_{j=1}^{q} a_{ji}\, Bel(\psi_i \wedge \phi) = f_i \;\middle|\; i = 1, \dots, s\right\}$,and to simplify the notation suppose that $\phi = \vee_{i=1}^{h} \alpha_i$. If $Bel^{(2)}$ satisfies K_2, then $Bel^{(2)}(|\emptyset)$ will satisfy K_2', $Bel^{(2)}(\emptyset|\emptyset) = 1$, $\left\{\sum_{j=1}^{r} a_{ji}\, Bel^{(2)}(\theta_i \wedge \phi) = Bel^{(2)}(\theta_i|\emptyset) = b_i \;\middle|\; i = 1, \dots, m\right\}$. Therefore, K_2' is consistent since K_2 is consistent by the assumption that $K_2 \in CL$. Similarly, K''_2 is consistent. Let $\rho^{\rightarrow} = H_q^L(K_2') = \tau^{\rightarrow} = H_q^L(K''_2)$, $\nu^{\rightarrow} = H_q^L(K_2)$,So $\sum_{i=1}^{h} \nu_i = c$. Now $Bel(\neg\emptyset) = 1 \Rightarrow \sum_{i=1}^{h} \tau_i = 0 = \sum_{i=h+1}^{J} \rho_i \Rightarrow \sum_{j=1}^{h} \tau_j = 1 = \sum_{i=h+1}^{J} \rho_j$. Also, $Bel(\neg\emptyset) = 1 \Rightarrow \sum_{j=1}^{h} \tau_j = 1 = \sum_{i=h+1}^{J} \rho_j$. $Bel(\emptyset) = c \Rightarrow \sum_{i=1}^{h} Bel(\alpha_i) = c$ from which it follows that

$$\sum_{i=1}^{J}(c\rho_i + (1-c)\tau_i) = c\sum_{i=1}^{J} \rho_i + (1-c)\sum_{i=1}^{h} \tau_i = c \qquad \text{(A.25)}$$

Hence we have $c\rho^{\rightarrow} + (1-c)\tau^{\rightarrow} \epsilon V^L(K_2)$. Similarly, we have $< \frac{\nu_1}{c}, \dots, \frac{\nu_h}{c}, 0,0,0 > \epsilon V^L(K_2')$ and $< 0,0,..,0, \frac{\nu_{h+1}}{1-c}, \dots, \frac{\nu_j}{1-c} > \epsilon V^L(K_2'')$. Now we shall study the following cases:

Case I: $\infty > q > 0$. The proof of case I is carried out by the definition of $H_q^L(K)$,when $\infty > q > 0$ and putting $E(x^{\rightarrow}) = \sum_{i=1}^{J} \nu_i^{q+1} = c^{q+1} E(\frac{\nu_1}{c}, \dots, \frac{\nu_h}{c}, 0,0,0) + (1-c)^{q+1} E(0,0,..,0, \frac{\nu_{h+1}}{1-c}, \dots, \frac{\nu_J}{1-c}) \geq E(c\rho^{\rightarrow} + (1-c)\tau^{\rightarrow}) \geq E(\nu^{\rightarrow})$ which implies, $\nu_i = c\rho_i$ for $i = 1, \dots, h$. So, we have $H_q^L(K_1)(\theta \wedge \phi) = H_q^L(K_2)(\theta \wedge \phi)$, for $\theta \epsilon SL$ and the result follows. The remaining cases for $q = 0$ and $-1 < q < 0$ are immediate.

A.6 The full proof of independence property

We shall consider the following: $K = \{Bel(p_1) = 1, Bel(p_2/p_1) = b = Bel(p_3/p_1)\}$

Now, we'll demonstrate that for $-1 < q, q \neq 0$, $H_q^L\ (K)(p_2 \wedge p_3|p_1) \neq b^2$.Here we shall study the following cases:

Case1: $q = 0$ **Case 2**: $-1 < q, q \neq 0$

Since $H_0^L = ME$,$H_q^L(K)$ satisfies independence for $q = 0$ by theorem (3.3.2). As for the remaining cases, assume $q > 0$ (the case for $-1 < q < 0$ is similar) and let $x_1 = Bel(\mathrm{p}_1 \wedge \mathrm{p}_2 \wedge \mathrm{p}_3), x_2 = Bel(\mathrm{p}_1 \wedge \mathrm{p}_2 \wedge \neg\mathrm{p}_3), x_3 = Bel(\mathrm{p}_1 \wedge \neg\mathrm{p}_2 \wedge \mathrm{p}_3), x_4 =$ $Bel(\mathrm{p}_1 \wedge \neg\mathrm{p}_2 \wedge \neg\mathrm{p}_3), x_5, x_6, x_7, x_8$ for the remaining atoms, we see that *K* yields the set of constraints

$$x_1 + x_2 + x_3 + x_4 = 1 \quad \text{(A.26)}$$

$$x_1 + x_3 = b \quad \text{(A.27)}$$

Implying

$$\mathrm{x}_3 = \mathrm{b}\text{-}\mathrm{x}_1 \quad \text{(A.28)}$$

And, we have

$$x_1 + x_2 = b \quad \text{(A.29)}$$

which gives

$$\mathrm{x}_2 = \mathrm{b}\text{-}\mathrm{x}_1 \quad \text{(A.30)}$$

Clearly, it follows that:

$$x_4 = 1 - 2b + x_1 \quad \text{(A.31)}$$

But $\sum_{i=1}^{8} x_i = 1$, together with (A.26) will give $x_5 + x_6 + x_7 + x_8 = 0$, i.e., that:

$$x_5 = x_6 = x_7 = x_8 = 0 \quad \text{(A.32)}$$

Therefore, we have

$$V^L(K) = \{< x_1, b - x_1, 1 - 2b + x_1, 0,0,0,0 > | 1 \geq x_1 \geq 0, 1 \geq 1 - 2b + x_1 \geq 0, 1 \geq b - x_1 \geq 0\} \quad \text{(A.33)}$$

Assuming $b < \frac{1}{2}$ and carrying on to find the range of x_1 in $V^L(K)$ which satisfies the above three inequalities we have $1 \geq b - x_1 \geq 0 \Rightarrow 1 - b \geq -x_1 \geq -b \Rightarrow b - 1 \leq x_1 \leq$

b.Since $b-1 \leq 0$, the condition $b-1 \leq x_1$ is implied by $0 \leq x_1$ and hence is redundant. So, we have

$$x_1 \leq b \tag{A.34}$$

Also, we have $1 \geq 1-2b+x_1 \geq 0$, which gives :

$$2b \geq x_1 \geq 2b-1 \tag{A.35}$$

But $b < \frac{1}{2}$ we have $2b-1 \leq 0$ and hence the condition $2b-1 \leq x_1$ is implied by $0 \leq x_1$, so it is redundant. Thus, we have:

$$x_1 \leq 2b \tag{A.36}$$

which is already implied by (A.34). Hence $V^L(K)$ will be $V^L(K) = \{< x_1, b-x_1, b-x_1, 1-2b+x_1, 0,0,0,0 > | x_1 \epsilon [0,b]\}$.Now we are after $x^{\rightarrow} \epsilon V^L(K)$ at which $f_q(x_1)$= $\sum_{i=1}^{J} x_i^{q+1}$ is minimum, $f_q(x_1) = x_1^{q+1} + 2(b-x_1)^{q+1} + (1-2b+x_1)^{q+1}$

$$f_q'(x_1) = (q+1)(x_1^q + 2(b-x_1)^q + (1-2b+x_1)^q$$

If $f_q'(x_1) = 0, \text{then}$

$$(x_1^q + 2(b-x_1)^q + (1-2b+x_1)^q) = 0 \tag{A.37}$$

We also have

$$f_q''(x_1) = q(q+1)(x_1^{q-1} + 2(b-x_1)^{q-1} + (1-2b+x_1)^{q-1} \tag{A.38}$$

$$f_q''(b^2) = q(q+1)(b^{q-1} + (1-b)^{q-1})^2 \tag{A.39}$$

Therefore, one gets $f_q''(b^2) > 0$ i.e., $x_1 = b^2$ will give a minimum value of $f_q < x_1, x_2, \dots, x_J >$. But $x_1 = b^2$ does not satisfy (A.38). To see this, assume that $x_1 = b^2$ satisfies (A.38). This implies:

$$(b^q)^2 - 2b^q(1-b)^q + ((1-b)^q)^2 = 0 \Leftrightarrow b = \frac{1}{2}$$

This is a contradiction. So, since $0 < b^2 < b$ (recall the range of x_1 here is$[0,b]$), this cannot not be a minimum point of $f_q(x_1^{\rightarrow})$.Therefore, we have $N^L(K)(p_2 \wedge p_3 | p_1) \neq$

$H_q^L(K)(p_2|p_1).H_q^L(K)(p_3|p_1) = b^2$(Since $H_q^L(K)(p_1) = 1$). Hence, H_q^L does not satisfy independence at $q = 0$.

Appendix B

The credibility of Rényian NME formalisms

Herewith, we are investigating the credibility of the derived Rényian closed form expression with respect to the four axioms of consistency.

B.1. Uniqueness

In axiomatic terms, uniqueness read as "If the same problem is solved twice in exactly the same way, the same answer is expected in both situations, i.e., the solution should be unique," according to this axiom (Shore 1980). As for Rényi's entropy (c.f., (4.6)), let the converse statement holds by assuming that:

$$\mathrm{H}_{\mathrm{q}}^{*}\left(\mathrm{f}_{\mathrm{q,N}}\right) = \mathrm{H}_{\mathrm{q}}^{*}\left(\mathrm{h}_{\mathrm{q,N}}\right) \text{ for distinct } \mathrm{f}_{\mathrm{q,N}},\ \mathrm{h}_{\mathrm{q,N}}, \mathrm{N} > 0 \qquad \text{(B.1)}$$

Hence, $\sum_{n=1}^{N} f_{q,N}^{q} = \sum_{n=1}^{N} h_{q,N}^{q}$, implying $\sum_{n=1}^{N}(f_{q,N}^{q} - h_{q,N}^{q}) = 0$. Additionally, assume the contrary, i.e., $f_{q,N}^{q} \neq h_{q,N}^{q}$.This directly implies the existence of positive real number $\gamma > 1$ satisfying

$$\mathrm{f}_{\mathrm{q,N}}^{\mathrm{q}} = \gamma\, \mathrm{h}_{\mathrm{q,N}}^{\mathrm{q}} \qquad \text{(B.2)}$$

Hence,

$$\frac{1}{1-q}\left(\ln\left(\sum_{n=1}^{N} f_{q,N}^{q}\right)\right) = \frac{1}{1-q}(\ln(\gamma)) + \frac{1}{1-q}\left(\ln\left(\sum_{n=1}^{N} h_{q,N}^{q}\right)\right) \qquad \text{(B.3)}$$

Hence, $f_{q,N}^{q} = h_{q,N}^{q}$, implies $\gamma = 1$, a contradiction.

Consequently, "the distinct $f_{q,N}, h_{q,N} \epsilon \Omega$ " can never share the same entropy functional. Hence , the axiom of uniqueness is satisfied by Rényi's NME formalism (see Shore 1980).

B.2. Invariance

In axiomatic terms, invariance reads as “The same solution should be obtained if the same inference problem is solved twice in two different coordinate systems” (see (Shore 1980)).Following (Shore 1980) approach, let Ξ be a coordinate transformation from state $\{S_n, n = 1,2, \ldots, N\}$ to state $\{M_n, n = 1,2, \ldots, N\}$, where M be a transformed set of N possible discrete states, namely $M = \{M_n, n = 1,2, \ldots, N\}$ with $\Gamma(p_{q,N}(M_n) = \Xi^{-1}(p_{q,N}(S_n)$, where J is the Jacobian$J = \frac{\partial(M_n)}{\partial(S_n)}$. In addition, assume that $\Gamma\Omega$ denotes the closed convex set of all probability distributions Γ defined on M satisfying $\Xi(p_{q,N}(M_n))$>0 for all $M_n \in M$, n = 1, 2, ...,N and $\sum_{n=1}^{N} \Xi\left(p_{q,N}(M_n)\right) = 1$. Hence, the Rényian information measure of (4.6) is invariant under transformations (Kayal 2017; Steinbrecher et al. 2016).

$$H_q^*\left(p_{q,N}\right) = H_q^*(\Xi\left(p_{q,N}\right)) \tag{B.4}$$

As a result of the correspondence between the minimality in Ω and $\Xi\Omega$, the axiom of invariance is fulfilled by EME closed form expression(Shore 1980).

B.3. System independence

The axiomatic definition of system independence (or, equivalently the additivity) is interpreted as “It should not matter whether one accounts for independent information about independent systems separately in terms of different probabilities or together in terms of the joint probability” (see (Shore 1980)). Eventually, the joint probability can be written as

$$h_{q,N}(x_k, y_n) = f_{q,N}(x_k)\, g_{q,N}(y_n), N > 0 \tag{B.5}$$

We can re-write Rényian functional(c.f., (4.6)) in the form:

$$H_q^*[h_{q,N}] = \frac{1}{1-q} \ln \left(\sum_k \sum_n h_{q,N}^q\right) \tag{B.6}$$

Engaging (B.5) and (B.6), one gets

$$\mathrm{H}_{\mathrm{q}}^{*}[\mathrm{f}_{\mathrm{q,N}}] + \mathrm{H}_{\mathrm{q}}^{*}[\mathrm{g}(\mathrm{Y}_{\mathrm{q,N}})] \neq \mathrm{H}_{\mathrm{q}}^{*}[\mathrm{h}_{\mathrm{q,N}}] \tag{B.7}$$

According to the inequality (B.7), "the joint EME state probability distribution of two independent non-extensive systems Q and V contradicts the axiom of system independence due to the presence of long-range interactions" (Shore 1980). This characteristic of the ME formalism is therefore obviously best suited to quantitatively analyse heavy queue-tailed with asymptotic power law behaviour dynamic systems of non-extensive information theoretic order q ($q \in (0.5,1)$ depicting long-range interactions as an inductive inference method.

It is worth noting that as non-extensive information theoretic order q reaches 1(Shannonian case) (B.7) changes to

$$\mathrm{H}_{1}^{*}[\mathrm{h}_{1,\mathrm{N}}] = \mathrm{H}_{1}^{*}[\mathrm{f}_{1,\mathrm{N}}] + \mathrm{H}_{1}^{*}(\mathrm{g}(\mathrm{Y}_{1,\mathrm{N}}) \tag{B.8}$$

As predicted, EME state probability distribution satisfies system independence, as shown by expression (B.8) (see Shore 1980). According to (Kouvatsos 2010), this is "a suitable property of EME formalism, as a method of inductive inference, for the study of short-range interactions extensive systems".

B.4. Subset independence

Subset independence can be interpreted (see Statistics 2007) by demonstrating that treating the entire density of the system is the same as investigating any independent subset of states of the underlying system from distinct conditional density perspective. Let ξ_{i} be the probability that a state of the system Q is in the set $\{S_i^*, i = 1,2,\dots,L\}$such that $\sum_i \xi_i = 1$.Moreover, let probability $\mathrm{f}_{\mathrm{q,i}}(\mathrm{x}_{\mathrm{ij}}) \epsilon\, \Omega_{\mathrm{i}}$, where Ω_{i}, is the closed convex set of all probability distributions on $\mathrm{S}_{\mathrm{i}}^{*}$, i.e., $\{\mathrm{f}_{\mathrm{q,i}}(\mathrm{x}_{\mathrm{ij}}) = \Pr\{\mathrm{X}_{\mathrm{i}} = \mathrm{x}_{\mathrm{ij}}\}\}$, where X_{i} is the state conditional random variable of the system $S_i^*, i = 1,2,\dots,L$. Furthermore, let x be an aggregate state of system Q and probability $\mathrm{f}_{\mathrm{q}}(x) \epsilon\, \Omega$, where X is the random variable defining the aggregate state of the system and $f_q(x) = \Pr\{X = x\}$ where X is the closed convex set of all probability distributions on S. Obviously, ξ_i is read as

$$\sum_{S_i^*} f_{q,i}(x_{ij}) = \xi_i \tag{B.9}$$

The total number of $\bigcup_{i=1}^{L} S_i^*$ states is employed to define the system Q' entire non-extensive entropy measure, $H_q^*(f_q)$. By employing the entropy measure of (4.6), we have

$$H_q^*(f_q) = \frac{1}{1-q} \ln\, (\sum_i \sum_{S_i} \xi_i\, f_{q,i}^q(x_{ij}) \tag{B.10}$$

where $f_q(x) \epsilon\, \Omega$. We can rewrite equation (B.10) as

$$H_q^*(f_q) = \frac{1}{1-q} \ln\, [\sum_i \xi_i \sum_{S_i} f_{q,i}^q(x_{ij})]\ ), , i = 1,2, \dots , L, \tag{B.11}$$

Let $H_i^*(f_{q,i})$ serves the conditional extensive entropy over the states S_i

$$H_{q,i}^*(f_{q,i}) = \frac{1}{1-q} \ln\, (\sum_{S_i^*} f_{q,i}^q(x_{ij})) \tag{B.12}$$

Thus, we have

$$\sum_{S_i^*} f_{q,i}^q(x_{ij}) = e^{(1-q)H_q^*(f_{q,i})} \tag{B.13}$$

Hence,

$$H_q^*(f_q) = \frac{1}{1-q} \ln\, (\sum_i \xi_i e^{(1-q)H_q^*(f_{q,i})})\ ) \tag{B.14}$$

The maximisation of $H_{q,i}^*(f_{q,i})$, on an individual basis, under constraints that are set conditionally, is identical to maximisation of $H_q^*(f_q)$, under accessible constraints . The assumption of subset independence is thus satisfied by the Rényi's ME formalism (see Shore 1980).

References

Aggarwal, V. (2005). “On Variable to Fixed length codes, Source Coding and Rényi’s Entropy.” Department of Electrical Engineering Indian Institute of Technology, Kanpur, India.

Agnew, CE. (1976). “Dynamic modelling and control of congestion–prone systems.” Operations Research 24(3):400–419.

Aguir, S., et al. (2004). “ The impact of retrials on call centre performance.” OR Spectrum 26(3):353–376.

Alfa, AS. (1979). “A numerical method for evaluating delay to a customer in a time inhomogeneous, single server queue with batch arrivals.” The Journal of the Operational Research Society 30(7):665–667.

Alnowibet, KA. and H. Perros.(2006). “ The nonstationary loss queue: a survey.” In: Barria JA, editor. Communication networks and computer systems—A tribute to professor Erol Gelenbe. Imperial College Press; p. 105–125.

Amari, S. (1985). “ Differential Geometrical Methods in Statistics.” Springer lecture Notes in Statistics 28.

Amari, S.(2001). “ *Information Geometry on Hierarchy of Probability Distributions*”. IEEE Transactions on Information Theory.

Amari, S.(2012). “Differential-geometrical methods in statistics” volume 28. Springer Science & Business Media.

Amigó, J.M,et al. (2018). “A Brief Review of Generalized Entropies.” Entropy 20(11), 813;https://doi.org/10.3390/e20110813

Anderson, E.J, et al. (2021). “Fundamental limits of bosonic broadcast channels.” In 2021 IEEE International Symposium on Information Theory (ISIT). (pp. 766-771). IEEE.

Arbabi, H., and I. Mezic. (2017). “ Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator.” SIAM Journal on Applied Dynamical Systems 16(4):2096-126.

Asokan, A. and Anitha, J. (2019). “Change detection techniques for remote sensing applications: a survey.” Earth Science Informatics12(2):143-60.

Avand, M. and Moradi, H. (2021). “Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability.” Journal of Hydrology 595, p.125663.

Baez,J.(2021). “ Information Geometry.” https://math.ucr.edu/home/baez/information/

Barbieri, A. L, et al. (2009). “An entropy-based approach to automatic image segmentation of satellite images.” Online available at: https://arxiv.org/PDF/0911.1759.PDF.

Basu ,et al.(1998). “Robust and efficient estimation by minimising a density power divergence.” Biometrika 85(3):549–559.

Beneš, V. E. (1965). “Mathematical Theory of Connecting Networks and Telephone Traffic.” Academic Press Inc.

Bertsimas, D. and G. Mourtzinou.(1997). “ Transient laws of non-stationary queueing systems and their applications.” Queueing Systems 25(1–4):115–155.

Boashash, B. and S.Ouelha. (2018). “Designing high-resolution time–frequency and time–scale distributions for the analysis and classification of non-stationary signals: a tutorial review with a comparison of features performance.” Digital Signal Processing 77:120-52.

Bratianu, C, et al. (2011). A. “Intergenerational learning dynamics in universities.” Electronic Journal of Knowledge Management 9(1), 10-18.

Bratianu, C. (2013).“Nonlinear integrators of the organizational intellectual capital.” In Fathi, M. (Ed.).Integration of practice-oriented knowledge technology: trends and perspectives pp. 3-17. Heidelberg: Springer

Bratianu, C.(2018). “Intellectual capital research and practice: 7 myths and one golden rule. Management & Marketing.” Challenges for the Knowledge Society 13(2), 859-879.

Bratianu, C. and R.Bejinaru. (2019). “The theory of knowledge fields: a thermodynamic approach.” Systems 7(2), 20, 1-12.

Bratianu, C. (2019). “Exploring knowledge entropy in organizations.” Management Dynamics in the Knowledge Economy 7(3):353-66.

Bratianu, C. and R.Bejinaru. (2020). “ Knowledge dynamics: a thermodynamic approach.”, Kybernetes 49(1), 6-21.

Brownlee, J. (2020). “Probability for Machine Learning: Discover How to Harness Probability with Python.”.

Burget, D. (2018).“ Rescaled entropy and cellular Automata.” Online available at: https://arxiv.org/PDF/2005.08585.PDF

Cantor, J.L. (1984). "Information Theoretic Analysis for a Multi-Server Queuing System of Equilibrium with Application to Queue in Tandem.", M. Sc. Thesis University of Maryland.

Cantor, J, et al. (1986). "Information theoretic analysis for a general queueing system at equilibrium with application to queues in tandem." Acta Informatica 23(6): 657-678.

Chakraborti, S, et al. (2021). "t-Entropy: A New Measure of Uncertainty with Some Applications.", arXiv:2105.00316v2

Chamorro, HR, et al. (2020). " Information Length Quantification and Forecasting of Power Systems Kinetic Energy." IEEE Transactions on Power Systems.

Chauhan, D. (2018). "Maximum entropy analysis of unreliable M^X/(G1,G2)/1 queue with Bernoulli vacation schedule." Int J Stat Appl Math 3(6):110-8.

Chen , G., et al. (2013). " Terminal appointment system design by non-stationary $M(t)/E_k/c(t)$queueing model and genetic algorithm." International Journal of Production Economics 146(2):694–703.

Choudhury, G.L., et al.(1997). "Numerical solution of piecewise stationary $M_t\,/G_t/1$ queues. Operations Research 45(3):451–463.

Chung, K.and D. Min. (2014). "Staffing a service system with appointment-based customer arrivals." , Journal of the Operational Research Society 65(10):1533–1543.

Cichocki et al.(2011) " Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization." Entropy13(1):134–170.

Cilingir, K, et al.(2020). "Deep Divergence Learning" Online arXiv:2005.02612v17

Clarke, A.B. (1956). "A waiting line process of Markov type." The Annals of Mathematical Statistics 27(2):452–459.

Cosmetatos, G.P. (1976). "Some approximate equilibrium results for the multi-server queue$(M/G/r)$." Operational Research Quarterly 27(3):615–620.

Cox, D. R. (2006). "Principles of Statistical Inference." Cambridge University Press.

Csiszár, I. (2008). "Axiomatic Characterizations of Information Measures." Entropy 10(3): 261-273.

Cunha, M.d.Vale, and C.C. Ribeiro Santos, et al. (2020). "Shannon entropy in time-varying semantic networks of titles of scientific paper." Appl Net Sci 5, 53. Online available at: https://doi.org/10.1007/s41109-020-00292-0

Cyr,S.and C.W.Choo.(2010).“The individual and social dynamics of knowledge sharing: an exploratory study.”, The Journal of Documentation 66(6), 824-846.

Czachórski, T., et al.(2009). “Diffusion approximation models for transient states and their application to priority queues.” International Journal on Advances in Networks and Services 2(2 and 3):205–217

Czachórski, T., et al.(2010). “A diffusion approximation model for wireless networks based on IEEE 802.11 standard.” Computer Communications 33:S86–S92.

Depeweg et al. (2016). “ Learning and policy search in stochastic dynamical systems with Bayesian neural networks.” arXiv preprint arXiv:1605.07127.

Dempster, A.P. (1997). “Upper and Lower Probabilities Induced by a Multi-valued Mapping.” Annals of Mathematical Statistics 38:325-339.

Deppoman, A. (2021). “Thermo-fractals, non-additive entropy, and q-calculus.”, Physics 3(2):290-301.

Dharmaprani, D, et al.(2021). “M/M/Infinity birth-death process- A quantitative representational framework to summarize and explain phase singularity and wavelet dynamics in atrial fibrillation.” Frontiers in physiology 11:616866

Di Crescenzo, A. and AG. Nobile. (1995). “Diffusion approximation to a queueing system with time-dependent arrival and service rates.” Queueing Systems 19 (1–2):41–62.

Dodson, C. T. J.(1999). “ Spatial Statistical and Information Geometry for Parametric Statistical Models of Galaxy Clustering.” Springer Netherlands 10.

Dodson, C.T.J. (2005). “Topics in Information Geometry.” University of Manchester. Available online on www.eprints.maths.manchester.ac.uk

Donglai, W, et al. (2018). “Learning and Using The Arrow of Time. ” CVF version and also available on IEEE Xplore. Online available at: https://openaccess.thecvf.com/content_cvpr_2018/papers/Wei_Learning_and_Using_CVPR_2018_paper.PDF

Duda, A. (1986). “Diffusion approximations for time-dependent queueing systems.” IEEE Journal on Selected Areas in Communications 4(6):905–918.

Eftimiadi,F. and EP. Trimigliozzi. (2019).“Reversible Computing: An Introduction.”OSF Preprints.

El-Affendi, M.A. and D.D.Kouvatsos. (1983). “A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems at equilibrium.” Acta Informatica 19, 339-

355.

Ershadi, M. J, et al. (2019). “Information Security Risk Management of Research Information Systems: A hybrid approach of Fuzzy FMEA, AHP, TOPSIS and Shannon Entropy.” J. Digit. Inf. Manag. 17.6 : 321.

Fang, S.C, et al.1997. “Entropy optimization models.” In Entropy Optimization and Mathematical Programming (pp. 17-49). Springer, Boston, MA.

Fisher, R.A.(1935). “The fiducial argument in statistical inference.” Annals of eugenics 6(4), pp.391-398.

Feldman, Z., et al.(2008). “Staffing of time-varying queues to achieve time-stable performance.” Management Science 54 (2):324–338.

Flandrin, P, et al. (2018). “Time-frequency reassignment: from principles to algorithms.” In Applications in time-frequency signal processing (pp. 179-204). CRC Press.

Flick, A., and M. Liao. (2010). “A queuing system with time varying rates.” Statistics & Probability Letters 80(5–6):386–389.

Frieden, B.R. (2004). “Science from Fisher information.” Citeseer vol.974.

Gabriel, L-G, et al. (2017).“Maximum Entropy Flow Networks.” ICLR. arXiv:1701.03504v2. Online available at: https://arxiv.org/PDF/1701.03504.PDF

Gao, J. and K.Ozbay. (2016). “Modeling double parking impacts on urban street.” In Transportation Research Board 95th Annual Mesting. Vol 12.

Ghosh, K. et al. (2017). “ A generalized divergence for statistical inference.” Bernoulli 23 (4A):2746–2783.

Ghosh,K.et al.(2016).“Robust bayes estimation using the density power divergence.” Annals of the Institute of Statistical Mathematics 68 (2):413–437.

Ghosh, K,et al. (2020). “The maximum caliber variational principle for nonequilibria.” Annual review of physical chemistry 20;71:213-38.

Giovannetti, V, et al. (2004). “Minimum Rényi and Wehrl entropies at the output of bosonic channels.” Massachusetts Institute of Technology – Research Laboratory of Electronics– Department of Mechanical Engineering 77 Massachusetts Ave., Cambridge, MA 02139, USA.

Goodfellow, I, et al. (2016). “Deep Learning (Adaptive Computation and Machine Learning Series).” Illustrated Edition.

Gomes, J.B., and MJ. Carneiro.(2021).“Polynomial Entropy for Interval Maps and Lap Number.” Qualitative Theory of Dynamical Systems 20(1):1-6.

González-Espinoza, A.et al.(2020). “Arrow of time across five centuries of classical music.” Physical Review Research 2(3):033166.

Green, L.V., and Nguyen, V.(2001). “ Strategies for cutting hospital beds: the impact on patient service.” Health Services Research,36: 421-442.

Green, L.V. (2003). “How many hospital beds? *Inquiry*.”, 39: 400-412.

Green, L.V.,et al.(2005). “ Using queueing theory to increase the effectiveness of physician staffing in the emergency department.” Academic Emergency Medicine.

Guel-Cortez, A.-J. And E. Kim. (2020). “Information length analysis of linear autonomous stochastic processes.” Entropy 22(11)

Guel-Cortez, A.-J. and E. Kim. (2021). “Information geometric theory in the prediction of abrupt changes in system dynamics.” Entropy vol. 23, no. 6, p. 694.

Guiasu, S.(1986). “Maximum entropy condition in queueing theory.” Journal of the Operational Research Society 37, 293-301.

Guichard, D. (2017). “Single and multivariable calculus: early transcendentals.” Available online at: https://www.whitman.edu/mathematics/multivariable/

Gunawardena, G.(2006). “Matrix Algebra for Beginners, part III, Department of Systems Biology.” Harvard Medical school MA02115, USA.

Hall, B.C. (2015). “The Matrix Exponential. In: Lie Groups, Lie Algebras, and Representations. ” Graduate Texts in Mathematics vol 222. Springer, Cham.

Hamasha, M, et al. (2016). “ Analysis of Service Delivery System Using Markov Chain Approach and M/G/1 Queuing Systems.” Asian Journal of Industrial Engineering. [Online] Available at: 67518-AJIE-KR (docsdrive.com)

Hampshire, RC., et al.(2009).“A time-varying call centre design via Lagrangian mechanics.” Probability in the Engineering and Informational Sciences 23(2):231–259.

Harrison P.G. and N.M.Patel.(1992). “Performance modelling of communication networks and computer architectures.”International Computer S. Addison-Wesley Longman Publishing Co., Inc.

Hawes, P. (2007). “Investigation of Properties of Some Inference Processes”. PhD Thesis University of Manchester.

Hern´andez-Lobato, et al. (2016).“Black-box α divergence minimization.” In International conference on machine learning (pp. 1511-1520). PMLR.

Heseltine, J. and E. Kim.(2016). “Novel mapping in non-equilibrium stochastic processes.”Journal of Physics A: Mathematical and Theoretical vol. 49, no. 17, p. 175002.

Heseltine, J., and E. Kim.(2019). “Comparing information metrics for a coupled Ornstein–Uhlenbeck process.” Entropy 21(8):775.

Heseltine,J. and E. Kim.(2019).“Comparing information metrics for a coupled Ornstein–Uhlenbeck process.” Entropy vol. 21, no. 8, p. 775.

Hollerbach,R, et al.(2018). “Information geometry of nonlinear stochastic systems.” Entropy: 20(8):550.

Hou , F. (2018). “Measuring knowledge for recognition and knowledge entropy.”, arXiv preprint arXiv:1811.06135.

Huang, Y, et al. (2018). “ The complexity and entropy analysis for service game model based on different expectations and optimal pricing.” Entropy 20(11):858.

Ingolfsson, A.,et al. (2007). “ A survey and experimental comparison of service-level-approximation methods for nonstationary $M(t)/M/s(t)$queueing systems with exhaustive discipline.” INFORMS Journal on Computing 19(2):201–214.

Jain, M, et al. (2021). “Supplementary variable technique (SVT) for non-Markovian single server queue with service interruption (QSI).” Operational Research 21(4):2203-46.

Jaynes, E.T. (1957) . “Information Theory and Statistical Mechanics.” The Physical Review, 106, 620-630. https://doi.org/10.1103/PhysRev.106.620

Jaynes, E. T. (1968). “Prior Probabilities.” IEEE Transactions on Systems Science and Cybernetics 4(3): 227-241.

Jaynes, E. T. (1978) “ Where do we Stand on Maximum Entropy? .”The Maximum Entropy Formalism Conference, M.I.T., USA, M.I.T. Press, Cambridge, MA, USA.

Jaynes, E. T. (2003). “Probability Theory : The Logic of Science.” NY, USA, Cambridge University Press.

Jefferson,J. (2018). “ Information Geometry.” [Online] available at: https://rojefferson.blog/page/2/

Jelinek, H.F. and R. Tuladhar, et al. (2021). “Diffusion Entropy vs. Multiscale and Rényi Entropy to Detect Progression of Autonomic Neuropathy.” Front.

Physiol. https://doi.org/10.3389/fphys.2020.607324
Jiroušek, R. and PP.Shenoy. (2018). "A new definition of entropy of belief functions in the Dempster–Shafer theory." International Journal of Approximate Reasoning 1;92:49-65.
Jizba, P. and J.Korbel. (2020). " When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics." Physical Review E(4):042126.
Jurdana, V, et al. (2021). "Sparse time-frequency distribution reconstruction based on the 2D Rényi entropy shrinkage algorithm." Digital Signal Processing 118:103225.
Kapur, J.N. (1986)."Four families of measures of entropy." Indian Journal of Pure and Applied Mathematics.
Kapur, J.N. (1993). "Maximum Entropy Models in Science and Engineering." John Wiley & Sons, Inc., New York.
Kayal, S., and S. Kumar.(2017). "Estimating Rényi entropy of several exponential distributions under an asymmetric loss function."REVSTAT-Statistical Journal 15(4), 501–522.
Ke, J.-C., and C.-H.Lin. (2008). "Maximum entropy approach for batch-arrival queue under N policy with an un-reliable server and single vacation." J. Comput. Appl. Math. 221(1): 1-15.
Kempa, W.M. and R.Marjasz. (2021)."Study on Transient Queue-Size Distribution in the Finite-Buffer Model with Batch Arrivals and Multiple Vacation Policy." Entropy 23, 1410. Online available at: https://doi.org/10.3390/e23111410.
Kimber RM., et al. (1977). " Marlow M, Hollis EM. Flow/delay relationships for major/minor priority junctions. " Traffic Engineering & Control 18(11):516–519.
Kimber RM. and EM. Hollis(1978). "Peak-period traffic delays at road junctions and other bottlenecks. "Traffic Engineering & Control 19(10):442–446.
Kim,E.(2016). "Geometric structure and geodesic in a solvable model of nonequilibrium process." Physical Review E vol. 93, no. 6, p. 062127. Entropy vol. 20, no. 8, p. 550.
Kim,E. and R. Hollerbach.(2017). "Signature of nonlinear damping in geometric structure of a nonequilibrium process." Physical Review E vol. 95, no. 2, p. 022137.

Kim, E.(2018). “Investigating information geometry in classical and quantum systems through information length.” Entropy vol. 20, no. 8.

Kim,E. and A.-J. Guel-Cortez. (2021). “Causal information rate.” Entropy vol. 23, no. 8, p. 1087.

Kim, E.(2021). “Information geometry and non-equilibrium thermodynamic relations in the over-damped stochastic processes.” Journal of Statistical Mechanics: Theory and Experiment vol. 2021, no. 9, p. 093406.

Korbel, J. et al. (2018). “Classification of complex systems by their sample-space scaling exponents.” New Journal of Physics 10;20(9):093007.

Kouvatsos, D.D. (1985). “Maximum Entropy Methods for General Queuing Networks.” In modelling Techniques and Tools for Performance Analysis (D. Potier. Eds.) North- Holland, Amsterdam, 589-608.

Kouvatsos, D.D. (1986). “A Maximum Entropy Queue Length Distribution for the G/G/1 Finite Capacity Queue.” Proc. of Perf., ACM Sigmet. 224-236.

Kouvatsos, D. D. (1986a). "Maximum entropy and the G/G/1/N queue." Acta Inf. 23(5): 545-565.

Kouvatsos, D. D. (1986b). “A maximum entropy queue length distribution for the G/G/1 finite capacity queue. ” Proceedings of the 1986 ACM SIGMETRICS joint international conference on Computer performance modelling, measurement and evaluation. Raleigh, North Carolina, USA, ACM: 224-236

Kouvatsos, D.D. (1988). “Maximum Entropy Analysis of the G/G/1 Queue at Equilibrium.” J. Opl. Res. Soc. Vol. 39,183-200.

Kouvatsos, D. D. and N.Tabet-Aouel. (1989). “A Maximum Entropy Priority Approximation for a Stable G/G/1 Queue.” Acta Inf. 27(3): 247-286.

Kouvatsos, D. D. and S. G. Denazis (1993). “Entropy maximised queueing networks with blocking and multiple job classes.” Performance Evaluation 17(3): 189-205.

Kouvatsos, D. D. (1994). "Entropy maximisation and queueing network models.", Annals of Operations Research 48(1): 63-126.

Kouvatsos, D. and N. Tabet-Aouel . (1994). “An ME-based approximation for multi-server queues with pre-emptive priority. ”European Journal of Operational Research 77(3): 496-515.

Kouvatsos, D. D, et al. (2003). “Performance modelling of GPRS with bursty multiclass traffic.” Computers and Digital Techniques, IEE Proceedings - 150(2): 75-85.

Kouvatsos, D.D. and S.A. Assi. (2010). “On the analysis of queues with heavy tails: a non-extensive maximum entropy formalism and a generalisation of the Zipf-Mandelbrot distribution.” In International Workshop on Performance Evaluation of Computer and Communication Systems Springer, 99–111.

Kouvatsos, D.D. and S.A. Assi. (2011). “Generalised entropy maximisation and queues with bursty and/or heavy tails”. In Network Performance Engineering Springer, 357–392.

Kouvatsos,D.D. and I.A.Mageed. (2021a). “Non-Extensive Maximum Entropy Formalisms and Inductive Inferences of Stable M/G/1 Queue with Heavy Tails.” in “Advanced Trends in Queueing Theory” Vol. 2, Vladimir Anisimov and Nikolaos Limnios (eds.), Books in ‘Mathematics and Statistics’, Sciences by ISTE & J. Wiley, London, UK.

Kouvatsos, D.D. et I.A. Mageed. (2021b). “Formalismes de maximum d’entropie non extensive et inférence inductive d’une file d’attente M/G/1 stable à queues Lourdes.” “Théorie des files d’attente 2”Théorie et Practique, Sous la direction de Vladimir Anisimov et Nikolaos Limnios, Mathématique, Sciences by ISTE & J. Wiley, Londres, Royaume-Uni.

Kolmogorov, A. (1931). “Sur le problème d'attente.” Matematicheskii Sbornik 38 (1–2):101–106 (in French).

Kondor, R. and S.Trivedi. (2018). “On the Generalization of Equivariance and Convolution in Neural Networks to the action of Compact Groups.” Proceedings of the 35th International Conference on Machine Learning Stockholm, Sweden, PMLR 80.

Kulkarni, V.G.(2016). “Modelling and analysis of stochastic systems.” Chapman and Hall/CRC.

Kumar, S., and A.Choudhary. (2011). “ A Noiseless Coding Theorem Connected with Generalized Rényi’s Entropy of Order α for Incomplete Power Probability Distribution p^{β}.” Asian Journal of Applied Sciences 4(6): 649-656,ISSN 1996-3343/ DOI: 10.3923/japs.2011.649.656 © 2011 Knowledge Review, Malaysia.

Kumar, B.K, et al.(2014). “Transient and steady-state analysis of queueing systems with catastrophes and impatient customers.” Int. J. Mathematics in Operational Research Vol. 6, No. 5.

Kybic, J. (2006). “Incremental updating of nearest neighbour-based high-dimensional entropy estimation." In Proc. Acoustics, Speech, and Signal Processing.

Landi, G.T. and M. Paternostro. (2021). "Irreversible entropy production: From classical to quantum." Rev.Mod.Phys 93, 035008. Online available at: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.93.035008

Lefter, V, et al. (2011). "Intergenerational knowledge transfer in the academic environment of the knowledge-based economy.", Amfiteatru Economic Journal 13(30), 392-403.

Leistner, F. (2010)."Mastering organizational flow: how to make knowledge sharing work.", New York, NY: John Wiley & Sons.

Lee, J.M. (1950). "Riemannian Manifolds: An Introduction to Curvature." Graduate Texts in Mathematics 176.Springer-Verlag New York, Inc.

Leonenko, N.N. (2010)."Statistical Inference for Shannon and Renyi Information." Recent Researches in Applied Mathematics,15th WSEAS International Conference on Applied Mathematics (MATH'10).Vouliagmeni, Athens, Greece.

Li Y. and RE. Turner. (2016). "Rényi divergence variational inference." Advances in neural information processing systems 29.

Lopez-Herrero, M. J. (2002). "On the number of customers served in the M/G/1 retrial queue: first moments and maximum entropy approach." Computers & Operations Research 29(12): 1739-1757.

Loveridge, L.C.(2016). "Physical and Geometric interpretations of the Riemannian Tensor, Ricci Tensor and Scalar Curvature." [Online] Available on https://arxiv.org/PDF/gr-qc/0401099.PDF

Lu, Z. (2017). "Functional Equations.''Lecture notes for the Math Circle, Irvine. Partially supported by the NSF grant DMS-1510232. Available Online at https://www.math.uci.edu/~mathcircle/materials/M6L2.PDF

Ly, A.(2017)."A tutorial on fisher information." Journal of Mathematical Psychology vol. 80, pp. 40–55.

Mageed, I.A. and D.D.Kouvatsos. (2010)."Extended Properties of the Class of Rényi Generalized Entropies in the Discrete Time Domain." The 12th Annual Post Graduate Symposium on the Convergence of Telecommunications Networking and Broadcasting (PGNet2010) Liverpool John Morse University, Available online at: http://www.cms.livjm.ac.uk/pgnet2010/MakeCD/index.htm

Mageed, I.A and D.D.Kouvatsos. (2011). "Extended Properties of the Class of Rényi

Generalized Entropies in the Discrete Time Domain.” Extended version of (Mageed and Kouvatsos 2010) IEEE *International Conference on Computer Networks and Information Technology* Abbottabad Pakistan. pp. 1-7, doi: 10.1109/ICCNIT.2011.6020894. Online available at https://ieeexplore.ieee.org/document/6020894

Mageed, I.A. and D.D.Kouvatsos. (2019).“Information Geometric Structure of Stable M/G/1 Queue Manifold and its Matrix Exponential.” Proceedings of the 35th UK Performance Engineering Workshop School of Computing University of Leeds, Edited by Karim Djemame, p.123-135. Available online at: https://sites.google.com/view/ukpew2019/home

Mageed, I.A. and D.D.Kouvatsos.(2021). “The Impact of Information Geometry on the Analysis of the Stable M/G/1 Queue Manifold.” Major extension of (Mageed and Kouvatsos 2019) In Proceedings of the 10th International Conference on Operations Research and Enterprise Systems - Volume 1: ICORES, ISBN 978-989-758-485-5, pages 153-160. DOI: 10.5220/0010206801530160.

Mandelbaum, A. and WA.Massey. “Strong approximations for time-dependent queues.” Mathematics of Operations Research 20(1):33–64.

Manish, K.H. (2020).“A Brief History of Entropy.” Towards Data Science. https://towardsdatascience.com/a-brief-history-of-entropy-chapter-1-9a2f1bc0d6de

Marvizadeh, S.Z.(2013).“Entropy Applications in Industrial Engineering.” PhD Thesis University of Nebraska – Lincoln.

Massey, WA. (2002). “The analysis of queues with time-varying rates for telecommunication models.” Telecommunication Systems 21(2–4):173–204.

Massey, WA. and J. Pender.(2013). “Gaussian skewness approximation for dynamic rate multi-server queues with abandonment.” Queueing Systems 75(2–4):243–277.

Mauro, R., and A.Pompigna.(2020). “State of the art and computational aspects of time-dependent waiting models for non-signalised intersections.” Journal of Traffic and Transportation Engineering (English Edition). 7(6):808-31.

Melin, J.,et al. (2020). “The role of entropy in construct specification equations (CSE) to improve the validity of memory tests.” Entropy 23(2):212.

Merriam, A. (2021). “A Tale of Genius with Tragedy — Ludwig Boltzmann.” Online available at:https://www.cantorsparadise.com/a-tale-of-genius-with-tragedy-ludwig-boltzmann-92c6a51dd387

Michalowicz, J.V,et al. (2013). “Handbook of differential entropy.” Crc Press.

Minka, T. (2005). “ Divergence measures and message passing.” Technical report Microsoft Research.

Montesinos, L, et al. (2018). “On the use of approximate entropy and sample entropy with centre of pressure time-series.” J Neuro Engineering Rehabil 15, 116. https://doi.org/10.1186/s12984-018-0465-9

Nanda, AK. and Chowdhury. (2021). “Shannon's Entropy and Its Generalisations Towards Statistical Inference in Last Seven Decades.” International Statistical Review 89(1):167-85.

Nakagawa,K.(2002). “ The geometry of M/D/1 queues on Large Deviation.” International Transactions of Operational Research Volume 9, Issue 2.

Nielsen, F. (2020). “ An Elementary Introduction to Information Geometry.”Sony Computer Science Laboratories Inc, Japan, MDPI Journal.

Nicholson, S,B. et al. “Time–information uncertainty relations in thermodynamics.” Nature Physics16.12 : 1211-1215.

Nonaka, I., and H. Takeuchi.(1995). “The knowledge-creating company. How Japanese companies create the dynamics of innovation.” Oxford, UK: Oxford University Press.

Norton, J.D. (2020). “ Non-Euclidean geometry and curved spaces.” Department of history and philosophy of science University of Pittsburgh, available online at https://www.pitt.edu/~jdnorton/teaching/HPS_0410/chapters/non_Euclid_curved/index.htm

Ollivier, Y.(2010). “A visual interpretation to Riemannian Curvatures and Some Discrete Generalizations.”[Online] Available on www.yann-ollivier.org/rech/publs/visualcurvature.PDF

Ovler, F.W, et al.(2010). “ Digital Library of Mathematical Functions.” National Institute of Standards and Technology from https://dlmf.nist.gov (release date 01-07-2011, Washington, DC. 2010)

Paluš, M,et al.(2020). “Causality, dynamical systems and the arrow of time.” Chaos: An Interdisciplinary Journal of Nonlinear Science28(7):075307.

Panta, AP,et al. (2021).“A Review of Vacation Queueing Models in Different Framework.” Annals of Pure and Applied Mathematics 24(2):99-121.

Parlar, M. (1984). “Optimal dynamic service rate control in time dependent $M/M/S/N$ queues.” International Journal of Systems Science 15(1):107–118.

Paris, J.B., and A.Vencovska.(1992). “A Method for Updating That Justifies Minimum Cross Entropy.” International Journal of Approximate Reasoning Vol 7, issue 1-2: 1-18, Elsevier Science Publishing Co. Inc.

Paris, J.B. (1994).“The Uncertain Reasoner’s Companion.” A Mathematical Perspective, Cambridge University Press .

Paris, J.B. (2003). “Uncertain Reasoning,” ASL, Helsinki.

Peg, L, et al., (2007). “The Geometric Structure of the Pareto Distribution.” Boleti′n de la Asociación Matemática Venezolana Vol. XIV, No 1 .

Pearl, J. (1990).“Reasoning with Belief Functions: An Analysis of Compatibility.” International Journal of Approximate Reasoning 4:363-389. Online available at https://ftp.cs.ucla.edu/pub/stat_ser/r136.PDF

Pishro-Nik, H. (2014). “Introduction to probability, statistics, and random processes.”, available at: https://wwww.p[robnabilitycourse.com, Kappa Research LLC.

Quah, J, et al.(2020). “Prospective cross-sectional study using Poisson renewal theory to study phase singularity formation and destruction rates in atrial fibrillation (RENEWAL-AF): Study design.” Journal of Arrhythmia 36(4):660-7.

Rachev, S.T.(1989). “ The problem of stability in queueing theory.” Queueing Syst 4, 287–317 Available online at: https://doi.org/10.1007/BF01159470

Rapisarda, A, et al. (2019). “Nonadditive Entropies and Complex Systems.” Entropy 21, 538. Available online on https://doi.org/10.3390/e21050538.

Regli, JB. and R. Silva. (2018). “Alpha-Beta Divergence For Variational Inference .”- arXiv preprint arXiv:1805.01045, - arxiv.org. Available online at [1805.01045] Alpha-Beta Divergence For Variational Inference (arxiv.org)

Rényi, A. (1961). “ On measures of entropy and information .” In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California.

Rex, A. (2017). “Maxwell’s Demon—A Historical Review.” Entropy 19, 240. Online at:https://doi.org/10.3390/e19060240.

Ribeiro, M, et al. (2021).“The entropy universe.” Entropy 23.2: 222.

Rödder,W. al. (2019). “Weight Prediction on Missing Links in Social Networks–a

Cross-Entropy-Based Approach.” Journal of Applied Logics—IfCoLog Journal of Logics and their Applications 6.1: 95-116.

Rudelius, T. (2012) “ A Geometric Understanding of Ricci Curvature in the context of Pseudo-Riemannian Manifold. ” Bachelor of Arts Cornell University.

Sason I. (2018). “Tight Bounds on the Rényi Entropy via Majorization with Applications to Guessing and Compression.” Entropy (Basel);20(12):896. doi: 10.3390/e20120896. PMID: 33266620; PMCID: PMC7512480.

Saulig, N, et al. Ž. (2017). “A local entropy-based algorithm for information content extraction from time–frequency distributions of noisy signals.” Digital Signal Processing 1;70:155-65.

Selinka, et al. (2016). “Time-dependent performance approximation of truck handling operations at an air cargo terminal. ” Computers & Operations Research 65:164–173.

Scharfenaker, E. and J.Yang. (2020). “Maximum entropy economics.” The European Physical Journal Special Topics 229(9):1577-90.

Schönherr, M. et al. (2022). “A theory of traces and the divergence theorem. ”Online available at: https://arxiv.org/PDF/2206.07941.PDF

Schwarz, JA, et al. (2016). “ Performance analysis of time-dependent queueing systems: Survey and classification.” Omega 63:170-89.

Shannon, C.E. (1948). “A mathematical theory of communication.” Bell System Technical Journal 27(3), 379–423.

Shore, J., and R. Johnson (1980). “Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy.” IEEE Transactions on Information Theory 26(1): 26-37.

Shore, J. E. (1982). “Information theoretic approximations for M/G/1 and G/G/1 queuing systems.” Acta Informatica 17(1): 43

Singh, AK, et al. (2021). “ Adaptive applications of maximum entropy principle.” In Progress in Advanced Computing and Intelligent Engineering (pp. 373-379). Springer, Singapore.

Smets,Ph.et al. (1991). “The transferable belief model.”: Lecture Notes Computer Science, Springer, Berlin/Heidlberg, “ Symbolic and Quantitative approaches to Uncertainty”. V 548/1991: 91-96, V 548/1991: 91-96, ISBN : 978-3540-54659-7

Statistics, T. (2007). “http://www.cscs.umich.edu/ Statistical mechanics for non-extensive systems and long-range interactions." notebooks. http://www.cscs.umich.edu/

Steinbrecher, G,et al. (2016) "Category theoretic properties of the A. Rényi and C. Tsallis entropies." Journal of Modern Physics 7, 251–266.

Stolletz, R.(2008 a). "Approximation of the non-stationary $M(t)/M(t)/c(t)$ –queue using stationary queueing models: The stationary backlog-carryover approach."European Journal of Operational Research190(2):478–493.

Stolletz, R. (2008 b). "Non-stationary delay analysis of runway systems." OR Spectrum 30(1):191–213.

Stolletz, R. and S. Lagershausen.(2013). " Time-dependent performance evaluation for loss waiting queues with arbitrary distributions." International Journal of Production Research 51(5):1366–1378.

Sudhindra, S, et al. (2017). "Knowledge transfer: an information theory perspective." Knowledge Management Research & Practice 15(3):400-12.

Tahmasebi, S. (2020). "Weighted extensions of generalized cumulative residual entropy and their applications." Communications in Statistics-Theory and Methods 49(21):5196-219.

Tang, X,et al. (2020). "Neural Membrane Mutual Coupling Characterization Using Entropy-Based Iterative Learning Identification." IEEE Access. Online available at:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9257495.

Tarabia, AMK.(2000). "Transient analysis of $M/M/1/N$ queue—an alternative approach.", Tamkang Journal of Science and Engineering 3(4):263–266.

Tarighi,B, et al. (2022). "Universal logarithmic correction to Rényi (Shannon) entropy in generic systems of critical quadratic fermions." Physical Review B 105(24):245109.

Teixeira, A. and L.Antunes. (2021). "On Conditional Tsallis Entropy." Entropy 23, 1427, https://doi.org/10.3390/e23111427

Theodore, M. (2022). "Links between entropy, complexity, and the technological singularity." Technological Forecasting and Social Change 176, 121457.

Thomas,R. (2015). "Kissing the curve – manifolds in many dimensions."Available Online:https://plus.maths.org/content/kissing-curve-manifolds-many-dimensions

Toomaj, A. and A. Di Crescenzo. (2020). "Connections between weighted generalized cumulative residual entropy and variance." Mathematics 8(7):1072.

Tsallis, C. (1988). “Possible generalization of Boltzmann-Gibbs statistics.” Journal of Statistical Physics 52 (1-2), 479–487

Tunstall, B.P. (1968). “Synthesis of noiseless compression codes,” Ph.D. dissertation. Georgia Inst. Technol., Atlanta.

Turner, R.E., and M.Sahani. (2011). “Two problems with variational expectation maximisation for timeseries models.” Bayesian Time series models pp. 115–138.

Uffink, J. (1995). “Can the Maximum Entropy Be Explained As a Consistency Requirement.” Department of History and Foundations of Mathematics and Science University of Utrecht, the Netherlands: 1-34.

Upton RA, Tripathi SK. (1982). “An approximate transient analysis of the $M_t/M/1$ queue.” Performance Evaluation 2(2):118–132.

Van de Coevering, MCT. (1995).“ Computing transient performance measures for the $M/M/1$ queue.” OR Spektrum 17:19–22.

Van Hulle, M.M.(2008). “Constrained subspace ICA based on mutual information optimization directly.” Neural Computation 20:964–973.

Van Erven, T. and Harremos, P. *(*2014*).* “Rényi divergence and Kullback-Leibler divergence.” IEEE Transactions on Information Theory 60(7), 3797-3820.

Walstra, R. J. (1985). “Nonexponential networks of queues: a maximum entropy analysis.”SIGMETRICS Perform. Eval. Rev. 13(2): 27-37.

Wang, K.-H, et al. (2002). “Maximum entropy analysis to the N policy M/G/1 queueing system with a removable server. ” Applied Mathematical Modelling 26(12): 1151.

Wang, D, et al. (2020). “ The sum of weighted normalized square envelope: A unified framework for kurtosis, negative entropy, Gini index and smoothness index for machine health monitoring.” Mechanical systems and signal processing1;140:106725.

Wenger, E. (1998). “Communities of practice: learning, meaning, and identity.” Cambridge, UK: Cambridge University Press.

Wijesuriya, KHK, et al. (2021). “Proof of Twin Prime Conjecture that can be obtained by using contradiction method in mathematics and it's applications to Thermal Physics.” https://osf.io/szce4

Worthington, D., and A. Wall. (1999). “Using the discrete time modelling approach to evaluate the time-dependent behaviour of queueing systems." Journal of the Operational Research Society 50(8):777–788.

Wu, J.-S., and Chan. W.C. (1989). "Maximum Entropy Analysis of Multiple-server Queueing Systems." Journal of the Operational Research Society 40(9): 815-825.

Xin, B,et al. (2020). " Exploration entropy for reinforcement learning.", Mathematical Problems in Engineering.

Xu, K., et al.(2014). "Time-varying performance analysis of multi-hop wireless networks with CBR traffic." IEEE Transactions on Vehicular Technology 63(7):3397–3409.

Yager, PR.(2018). "Interval valued entropies for Dempster–Shafer structures". Knowledge-Based Systems 161: 390-397.

Yang, D. Y,et al. (2011). "First two moment entropy maximisation approach for M/G/1 queues with second optional service and server breakdowns." International Journal of Services Operations and Informatics 6(4): 310-331.

Yom-Tov, GB. and A.Mandelbaum.(2014). "Erlang-R: a time-varying queue with re-entrant customers, in support of healthcare staffing." Manufacturing & Service Operations Management 16(2):283–299.

Zhang, B. and P.Shang.(2020)."Measuring information transfer by dispersion transfer entropy." Communications in Nonlinear Science and Numerical Simulation Vol 89, https://doi.org/10.1016/j.cnsns.2020.105329

Zhang, Q, et al. (2021). " Non-Markovianity and the Landauer principle in composite thermal environments." Physical Review A103(3):032201.

Zhang, X. (2020). "Maximum entropy distribution with fractional moments for reliability analysis." Structural Safety 1;83:101904.

Zhao,X. and P. Shang.(2015). "Measuring the Uncertainty of Coupling." A Letter Journal of the Frontiers of Physics.

Zegers, P.(2015). "Fisher information properties." Entropy vol. 17, no. 7, pp. 4918–4939.

Zitnick, C.L. (2003). "Computing Probabilities in Large Domains by Maximizing Rényi's Quadratic Entropy." PhD Thesis, CMU-RI-TR-03-20, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15217.

Zunino, L. and C.W. Kulp.(2017). "Detecting nonlinearity in short and noisy time series using the permutation entropy." Physics Letters A13;381(42):3627-35.