Advanced Maximum Entropy Approaches for Medical and Microscopy Imaging

Zahra Amini Farsan

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität, München

vorgelegt von
Zahra Amini Farsan

München, den 5. Mai 2021

Erster Berichterstatter: **Prof. Dr. Volker J. Schmid**
Zweiter Berichterstatter: **Prof. Liqun Wang**
Dritter Berichterstatter: **Prof. Dr. Thomas Augustin**

Tag der Einreichung: 5. Mai 2021
Tag der Disputation: 26. Juli 2021

Acknowledgment

My thesis would not have been possible to complete without the help and support of many people:

At first, I would like to express my special appreciation and thanks to my supervisor Prof. Dr. Volker J Schmid, you have been a tremendous supervisor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career has been priceless. I would also like to thank my committee members, Prof. Dr. Liqun Wang, Prof. Dr. Thomas Augustin, Prof. Dr. Christian Heumann and Prof. Dr. Göran Kauermann for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you. All of you have been there to support me when I recruited patients and collected data for my Ph.D. thesis.

Thank you also for the financial support of the LMU-Statistics institute, which gave me the chance to present my work at several international conferences. Thank you also to all people who helped me to improve this thesis by carefully proofreading parts of it.

A special thanks to my parents. Words cannot express how grateful I am to my mother and father for all the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far.

I would also like to thank all of my friends who supported me in writing and incented me to strive towards my goal. In the end, I would like express appreciation to my beloved husband Dr. Majid Sharifipour who spent sleepless nights with and was always my support in the moments when there was no one to answer my queries.

Finally, I could not have completed this dissertation without the support of my friend, Dr. Christina Schneider, who provided stimulating discussions as well as happy distractions to rest my mind outside my research.

There shall be wisdom, Knowledge and decency. "Ferdowsi"

Zusammenfassung

Die Methode der maximalen Entropie ist ein wichtiger Bestandteil der statistischen Inferenz, die in immer stärkerem Maße für die Konstruktion von Modellen verwendet wird, die biologische Systeme, insbesondere komplexe Systeme, aus empirischen Datensätzen beschreiben und vorhersagen können. In diesen ertragreichen Anwendungen ist es von besonderem Interesse, exakte Verteilungsfunktionen mit minimaler Information über die Eigenschaften der Daten und ohne Ausnutzung menschlicher Subjektivität zu bestimmen. In dieser Arbeit wird durch eine Kombination der Maximum-Entropie-Methode mit geeigneten Optimierungsverfahren ein automatisiertes Verfahren verwendet, um dieses Ziel für univariate und bivariate Daten zu erreichen. Notwendige Eigenschaften von Zufallsvariablen sind lediglich ihre Stetigkeit und ihre Approximierbarkeit als unabhängige und identisch verteilte Variablen. In dieser Arbeit versuchen wir, zwei numerische probabilistische Algorithmen präzise zu präsentieren und sie zur Schätzung der univariaten und bivariaten Modelle der zur Verfügung stehenden Daten anzuwenden.

Zunächst wird mit einer Kombination aus der Maximum-Entropie Methode, der Newton-Methode und dem Bayes'schen Maximum-A-Posteriori-Ansatz die Schätzung der kinetischen Parameter mit arteriellen Eingangsfunktionen (AIFs) in Fällen ohne Messung der AIF ermöglicht. Die Ergebnisse zeigen, dass die AIF aus den Daten der dynamischen kontrastverstärkten Magnetresonanztomographie (DCE-MRT) mit der Maximum-Entropie-Methode zuverlässig bestimmt werden kann. Anschließend können die kinetischen Parameter gewonnen werden. Durch die Anwendung der entwickelten Methode wird eine gute Datenanpassung und damit eine genauere Vorhersage der kinetischen Parameter erreicht, was wiederum zu einer zuverlässigeren Anwendung der DCE-MRT führt.

Im bivariaten Fall betrachten wir die Kolokalisierung zur quantitativen Analyse in der Fluoreszenzmikroskopie-Bildgebung. Die in diesem Fall vorgeschlagene Methode ergibt sich aus der Kombination der Maximum-Entropie-Methode (MEM) und einer Gaußschen Copula, die wir Maximum-Entropie-Copula (MEC) nennen. Mit dieser neuartigen Methode kann die räumliche und nichtlineare Korrelation von Signalen gemessen werden, um die Kolokalisierung von Markern in Bildern der Fluoreszenzmikroskopie zu erhalten. Das Ergebnis zeigt, dass MEC in der Lage ist, die Ko- und Antikolokalisation auch in Situationen mit hohem Grundrauschen zu bestimmen. Der wesentliche Punkt hierbei ist, dass die Bestimmung der gemeinsamen Verteilung über ihre Marginale ein entscheidendes inverses Problem ist, das eine mögliche eindeutige Lösung im Falle der Wahl einer geeigneten Copula gemäß dem Satz von Sklar hat. Diese neu entwickelte Kombination aus Gaußscher Kopula und der univariaten

Maximum Entropie Randverteilung ermöglicht die Bestimmung einer eindeutigen bivariaten Verteilung. Daher kann ein Kolokalisationsparameter über Kendall's t ermittelt werden, der üblicherweise in der Copula-Literatur verwendet wird.

Die Bedeutung der Anwendung dieser Algorithmen auf biologische Daten lässt sich im Allgemeinen mit hoher Genauigkeit, schnellerer Rechengesch windigkeit und geringeren Kosten im Vergleich zu anderen Lösungen begründen. Die umfassende Anwendung und der Erfolg dieser Algorithmen in verschiedenen Kontexten hängen von ihrer konzeptionellen Eindeutigkeit und mathematischen Gültigkeit ab.

Anschließend wird eine Wahrscheinlichkeitsdichte durch iterative Erweiterung von kumulativen Verteilungsfunktionen geschätzt, wobei die geeignetsten Schätzungen mit einer Scoring-Funktion quantifiziert werden, um unregelmäßige Schwankungen zu erkennen. Dieses Kriterium verhindert eine Unter- oder Überanpassung der Daten als Alternative zur Verwendung des Bayes-Kriteriums. Die durch statistische Schwankungen in Stichproben induzierte Unsicherheit wird durch mehrfache Schätzungen für die Wahrscheinlichkeitsdichte berücksichtigt. Zusätzlich werden als nützliche Diagnostik zur Visualisierung der Qualität der geschätzten Wahrscheinlichkeitsdichten skalierte Quantil-Residuen-Diagramme eingeführt. Die Kullback-Leibler-Divergenz ist ein geeignetes Maß, um die Konvergenz der Schätzungen für die Wahrscheinlichkeitsdichtefunktion (PDF) zu der tatsächlichen PDF als Stichprobe anzuzeigen. Die Ergebnisse zeigen die generelle Anwendbarkeit dieser Methode für statistische Inferenz mit hohem Ertrag.

Summary

The maximum entropy framework is a cornerstone of statistical inference, which is employed at a growing rate for constructing models capable of describing and predicting biological systems, particularly complex ones, from empirical datasets. In these high-yield applications, determining exact probability distribution functions with only minimal information about data characteristics and without utilizing human subjectivity is of particular interest. In this thesis, an automated procedure of this kind for univariate and bivariate data is employed to reach this objective through combining the maximum entropy method with an appropriate optimization method. The only necessary characteristics of random variables are their continuousness and ability to be approximated as independent and identically distributed. In this work, we try to concisely present two numerical probabilistic algorithms and apply them to estimate the univariate and bivariate models of the available data.

In the first case, a combination of the maximum entropy method, Newton's method, and the Bayesian maximum a posteriori approach leads to the estimation of the kinetic parameters with arterial input functions (AIFs) in cases without any measurement of the AIF. The results shows that the AIF can reliably be determined from the data of dynamic contrast enhanced-magnetic resonance imaging (DCE-MRI) by maximum entropy method. Then, kinetic parameters can be obtained. By using the developed method, a good data fitting and thus a more accurate prediction of the kinetic parameters are achieved, which, in turn, leads to a more reliable application of DCE-MRI.

In the bivariate case, we consider colocalization as a quantitative analysis in fluorescence microscopy imaging. The method proposed in this case is obtained by combining the Maximum Entropy Method (MEM) and a Gaussian Copula, which we call the Maximum Entropy Copula (MEC). This novel method is capable of measuring the spatial and nonlinear correlation of signals to obtain the colocalization of markers in fluorescence microscopy images. Based on the results, MEC is able to specify co- and anti-colocalization even in high-background situations. The main point here is that determining the joint distribution via its marginals is an important inverse problem which has one possible unique solution in case of choosing an proper copula according to Sklar's theorem. This developed combination of Gaussian copula and the univariate maximum entropy marginal distribution enables the determination of a unique bivariate distribution. Therefore, a colocalization parameter can be obtained via Kendall's t, which is commonly employed in the copula literature.

In general, the importance of applying these algorithms to biological data is attributed to the higher accuracy, faster computing rate, and lower cost of solutions in comparison to those of others. The extensive application and success of these algorithms

in various contexts depend on their conceptual plainness and mathematical validity.

Afterward, a probability density is estimated via enhancing trial cumulative distribution functions iteratively, in which more appropriate estimations are quantified using a scoring function that recognizes irregular fluctuations. This criterion resists under and over fitting data as an alternative to employing the Bayesian criterion. Uncertainty induced by statistical fluctuations in random samples is reflected by multiple estimates for the probability density. In addition, as a useful diagnostic for visualizing the quality of the estimated probability densities, scaled quantile residual plots are introduced. Kullback–Leibler divergence is an appropriate measure to indicate the convergence of estimations for the probability density function (PDF) to the actual PDF as sample. The findings indicate the general applicability of this method to high-yield statistical inference.

Contents

Chapter 1

Introduction

Determining the probability density of a random variable based on observations is a major and old issue in statistics. In recent years, various parametric and non-parametric methods have been introduced for determining the class of different statistical distributions. A standard way to estimate an unknown density is to identify its associated properties, such as symmetry, mode, and amplitude, so that choosing a distribution that applies to these properties and estimating its parameters let us approach the main goal. Deciding the best estimator is one challenge for estimating an unknown density. Researchers have introduced various methods for determining the statistical distribution and probability density, including the least squares method of mean error (Tuchler et al., 2002), Bayesian least squares method (Raphan and Simoncelli, 2007), and Best Linear Unbiased Estimator (Chow and Lin, 1976).

In general, in the case of continuous random variables, a probability density function (PDF) assigns a probability for the observation of a value falling within a specific given range. Empirically determining a PDF corresponding to N samples of univariate data has been investigated extensively in mathematics, with ubiquitous significance for practical applications. Multiple estimation approaches have been used with success for fitting a random data sample to parameters of a known functional form. Nevertheless, the functional form describing the underlying random process is often unknown.

In such cases, a specific functional form is usually regarded for convenience, particularly when data are limited. A combined model developed by a linear superposition of known functional forms is commonly employed in cases where the data have particular characteristics. Proceeding in all these situations requires expert knowledge. In a more general sense, for estimating a PDF in case of not knowing a parametric functional form for the PDF, nonparametric techniques are also available. For instance,for high-yield applications, including those in finance and bioinformatics, determining exact probability distribution functions with minimum data regarding data features and without employing human subjectivity is of particular interest (Farmer and Jacobs, 2018).

For that, the maximum entropy method is used today as a major method for estimating and determining the probability density with high accuracy and efficiency and, low bias. This method is employed to obtain the unknown density by solving optimization problems and is regarded as one of the most efficient methods capable of yielding maximum possible information for unknown density using the limited and known available information (Ebrahimi et al., 2008; Pougaza and Djafari, 2011; Thomas

and Cover, 2006). To reach this objective, such an automated technique for univariate data is applied via combining the maximum entropy method with single order statistics and maximum likelihood. Random variables just need to be continuous, to have the ability to be approximated as independent and identically distributed Farmer and Jacobs (2018).

The moment-constrained maximum entropy problem provides an estimation of a probability density with the greatest uncertainty among all densities meeting the supplied moment constraints. This issue appears in various fields including solid state physics (Haydock and Nex, 1984, 1985), Economy (Ormoneit and White, 1999), geophysical applications especially weather forecasting (Abramov and Majda, 2004; Haven et al., 2005; Kleeman, 2002), Renewable Energy Systems (Moreno and García-Álvarez, 2011; Wang and Zabaras, 2004), Electrical Engineering - Control and Power (Haddad et al., 2007), Image Processing in Medical Sciences (Mohammad-Djafari, 1992), biology (De Martino and De Martino, 2018) and in the other fields (Djafari and Demoment, 1990; Golan et al., 1997; Pougaza and Djafari, 2011; Soize, 2008).

The probability density itself is approximated via maximizing the Shannon entropy under constraints that are provided by the measured moments (phase-space-averaged monomials of the problem variables) (Mead and Papanicolaou, 1984). By using a standard procedure, (Wu et al., 2001) the constrained maximum entropy problem is transformed into the unconstrained minimization problem with the dual objective function. For further discussion about the theoretical basis of the maximum entropy moment problem, see (Borwein, 1991; Frontini and Tagliani, 1994; Fuglede, 1983; Tagliani, 1999). Recently, new algorithms for the multidimensional moment-constrained maximum entropy problem have been proposed in (Abramov, 2006, 2007, 2009; Abramov et al., 2010). Although the approach presented in Abramov (2006) is relatively simple and only applies to two-dimensional maximum entropy problems with moments of orders up to 4, the enhanced algorithm in Abramov (2007) employs an appropriate orthonormal polynomial base in the space of Lagrange multipliers for improving the convergence of its iterative optimization process.

It can practically solve 2D problems with up to 8th-order moments, 3D problems with up to 6th-order moments, and 4D problems up to 4th-order moments, adding up to 44, 83 and 69 moment constraints, respectively, without considering the normalization constraint for a probability density. Abramov (2009) introduced additional improvements to the algorithm for the multidimensional moment-constrained maximum entropy problem, including several Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterations (Broyden, 1970; Byrd et al., 1995; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) to advance between polynomial reorthonormalization points in an adaptive manner, versus single Newton steps proposed in (Abramov, 2007), and appropriate constraint rescaling to lower the magnitude differential between high- and low-order moment constraints.

The maximum entropy framework is a cornerstone of statistical inference, which is employed at a growing rate for constructing models capable of describing and predicting biological systems, particularly complex one, from empirical datasets. Entropy maximization or related concepts have been frequently utilized in the past ten years to analyze large biological datasets in various fields.

These fields range from determining macromolecular interactions and structures

(Boomsma et al., 2014; Cavalli et al., 2013; Cesari et al., 2018; Cofré et al., 2019; D'haeseleer et al., 2000; Ekeberg et al., 2013; Farré and Emberly, 2018; Hopf et al., 2012; Jaynes, 2003; Jennings et al., 2020; Ozer, 2008; Pitera and Chodera, 2012; Roux and Weare, 2013; Seno et al., 2008; Weigt et al., 2009; Zhang and Wolynes, 2015) to inferring signaling (Dhadialla et al., 2009; Lezon et al., 2006; Locasale and Wolf-Yadlin, 2009; Remacle et al., 2010; Sanguinetti et al., 2019) and regulatory networks (Graeber et al., 2010; Schneidman et al., 2006; Sharan and Karp, 2013; Shin et al., 2011) and the coding organization in neural populations (Cocco et al., 2009; Ferrari et al., 2017; Granot-Atedgi et al., 2013; Mora et al., 2010; Nghiem et al., 2018; Ohiorhenuan et al., 2010; Quadeer et al., 2020; Rostami et al., 2017; Roudi et al., 2009; Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Tkačik et al., 2013, 2010; Yeh et al., 2010; Yeo and Burge, 2004) based on DNA sequence analyses (the detection of specific binding sites, for instance) (Fariselli et al., 2020; Fernandez-de Cossio-Diaz and Mulet, 2019; Mora et al., 2010; Santolini et al., 2014; Yeo and Burge, 2004).

In addition, the more specialized features of the maximum entropy inference from the standpoint of statistical physics, computational biology, or information theory have been covered in multiple high-quality reviews, including the very recent ones (see e.g. De Martino and De Martino (2018); Jaynes (1957a,b); Lesne (2014); Pressé et al. (2013); Stein et al. (2015)).

Here, a concise and basic introduction to entropy maximization and its applicability for deriving models from biological datasets, especially in medical sciences and image processing is provided. We start from the basics (the concept of entropy and image processing) and end with a recent application of this method (a maximum entropy study of cellular metabolism functions). Our main goal is to specify its many applications and ability to provide new biological knowledge with minimum mathematical descriptions. However, to supplement the provided information, several fundamental mathematical concepts are inevitably provided in the Supplementary Material. Although this field is vast in scope and has been accompanied by many elegance since the 1940s, we, however, focus on the aspects that are of greater immediate relevance to our purposes. What follows is a short list of further components and new directions. For further information, the reader can refer to the suggested literature.

In line with our goals for entropy maximization, this obvious idea must be considered first: in deriving a statistical model from data, introducing biases other than those already existing in the data must be avoided, since these biases would be discretionary and unwarranted. As an example, when modeling a process with E possible outcomes (e.g. throwing a dice) without prior knowledge of it, the best estimate for a probability law governing this process would have to be the uniform distribution, in which each outcome has a probability (De Martino and De Martino, 2018).

Essentially, the entropy maximization framework extends this idea to cases with greater complexity and presents a guideline for constructing the optimal (least biased) probability distribution consistent with a specified set of constraints derived from data. Surely, the notion of entropy is central in this respect.

Hence, the main goal of this study is to investigate different applications of the maximum entropy method in the medical sciences. Especially, we focused on the cell biology and image processing. To approach the overall research objectives, five detailed

applications were selected. A brief description of the general framework of thesis and each application is as follows. The first and the second chapters includes introduction, literature review, and basic definitions about entropy and image processing.

In the first application, the problem was as follows. When deriving physiological kinetic parameters from Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) data, determining the arterial input function (AIF) is of particular interest. In this study, a Bayesian approach is proposed to capture the physiological parameters of DCE-MRI together with the AIF in cases with no measurement of the AIF. The proposed algorithm is obtained by combining the maximum entropy method (MEM) with the maximum a posteriori (MAP) method and is considered as an alternative for assessing the input function from the available data. By using the developed method, a good data fitting and thus a more accurate prediction of the kinetic parameters are achieved, which, in turn, leads to a more reliable application of DCE-MRI (chapter 3, Farsani and Schmid, 2017).

In the second application, the issue involved determining the joint probability distribution from its marginals considering the correlation between them. Since determining a joint probability distribution from its marginals, is regarded as an important problem in statistics, we linked this problem to an important problem in cell biology, which is to find the relation and location of the dye proteins in the cell and to reconstruct an image from its projections. A widely used technique to quantitatively analyze fluorescence microscopy images is the co-localization analysis. By localizing marked proteins in the nucleus of cells, a deep knowledge about biological processes in the nucleus is obtained. Many criteria have been proposed for measuring the co-localization of two markers; however, these criteria rely upon threshold the background subjectively and the assuming linearity. A robust method was proposed for capturing the bivariate distribution function of two color channels so that their co- or anti-colocalization can be quantified. The maximum entropy method (MEM) and a Gaussian copula were combined to develop this method, which is termed the maximum entropy copula (MEC). This novel method is capable of measuring the non-linear and spatial association of signals to specify the colocalization of markers in fluorescence microscopy images (chapter 4, Farsani and Schmid, 2021).

In the third application, we modified the previous algorithm in (Farsani and Schmid, 2017) the kinetic models used in contrast-based medical imaging, in which determining the arterial input function (AIF) is essential for the estimation of physiological parameters of the tissue via solving a nonlinear inverse problem named. Therefore, we estimate the AIF based on the modified maximum entropy method. The modified algorithm is obtained by combining the maximum entropy method and an optimization method, which was termed the modified maximum entropy method (MMEM). Then, we applied this algorithm in a Bayesian framework to estimate the kinetic parameters via the unique form of the arterial input function (chapter 5, submitted to Entropy).

In the fourth application, we considered the colocalization analysis. In this analysis, it was determined whether sub-cellular structures are located at the same physical position in which they can interact with each other. For an image pair, a green channel is typically analyzed with a red channel to specify their colocalization pixel. We

connected this problem to an ill-posed inverse problem in statistics to obtain the joint probability distribution from its marginals considering the concept of copula. Here, different classes of copula are considered to reconstruct the joint probability distribution with the marginals and evaluate which of them measures correlation more accurately. The proposed models are categorized into three groups based on the class of copula, the univariate maximum entropy model of both channels and their model structures. The generalized exponential distribution has been proposed as an appropriate model to describe the data. The focus of the current study is on what we can learn from the parameters of the distribution about the signals. We sell point for the article compared to the previous paper and look into the parameters to gain further insight into structure of the signal. Then, a numerical example is presented to illustrate the formulation and implementation of each type of the entropy copula model. In addition, the potential application of the maximum entropy copula in the fluorescence microscopy images data are discussed (chapter 6).

The thesis concludes with a short summary and an outlook to potential aspects of future research in chapter 7.

Software

All computations were carried out using the mathematical software MATLAB.

Chapter 2

Basic Concepts and Definitions of Entropy and Image Processing

2.1 Introduction

Information theory is based on statistics and probability with Shannon pioneering research that was founded as a mathematical theory of communication. A key indicator in information theory is Shannon's entropy, which is the core of information theory. Entropy is a measure of the uncertainty of a system, meaning that the system is more uncertain systems with less uncertainty, has been less reliable. Uncertainty is sometimes called a state of limited knowledge or "doubt." Something is uncertain if it is possible but not known. On the other hand, a proposition is uncertain if it is consistent with knowledge but not implied by knowledge. Connecting this notion of uncertainty with that one of information means that uncertainty is the amount of expected information that observations or experiments could reveal.

Therefore, it is important to note that the views of information and its relationship with uncertainty expressed above are common to all researchers analyzing random processes or random data. But, for information that is absolutely certain, what is the link between information and probability (or a degree of rational belief) or uncertainty?

In 1871, Boltzmann extends the subject of thermodynamics, by introducing probabilities (Boltzmann, 1871). Many authors have tried to sort out what Boltzmann thought and when he thought it (Bryan, 1894). With this in mind, we attempt to summarize Boltzmann's contribution to entropy with modern notation (Thomas and Cover, 2006).

2.2 Concept of Entropy

Entropy is expected information. It reflects what we expect to learn from observations, on average, and it depends on how we measure information. Entropy can be viewed as a measure of uniformity. Similarly, but within a different context, entropy is also a measure of disorder of a system. The second law of thermodynamics states that the entropy of a (closed) system (like the universe) increases with time. It represents the progression of the system toward equilibrium that is reached at the highest level of entropy. In more technical words, entropy is a measure of uncertainty of a single random variable (in statistics). The entropy of a random variable is a measure of the uncertainty of the random variable; It is a measure of the amount of information required on average to describe the random variable. In 1948, (Shannon, 1948b) introduced the entropy for both discrete and continuous random variables, based on Boltzmann's theory (Thomas and Cover, 2006).

We first introduce the concept of entropy, which is a measure of the uncertainty of a random variable. Let X be a discrete random variable with alphabet X and probability mass function $p(x) = Pr\{X = x\}, x \in \chi$. We denote the probability mass function by $p(x)$rather than $p_X(x)$, for convenience. Thus, $p(x)$ and $p(y)$ refer to two different random variables and are in fact different probability mass functions, $p_X(x)$ and $p_Y(y)$, respectively (Thomas and Cover, 2006).

2.2.1 Shannon's Entropy

Let X be a discrete random variable with alphabet χ and probability mass function $p(x) = p_r(X = x)$, entropy $H(X)$ is defined as follows:

$$H(X) = -\sum_{x\in\chi} p(x)logp(x), \tag{2.1}$$

The entropy of X can also be interpreted as the expected value of the random variable $log\frac{1}{p(X)}$, where X is drawn according to probability mass function $p(x)$. Thus,

$$H(X) = E_p(log\frac{1}{p(x)}) \tag{2.2}$$

Lemma1– $H(X) \geq 0$.
Proof– $0 \leq p(x) \leq 1$ implies that $log\frac{1}{p(x)} \geq 0$
Lemma2– $H_b(X) = (logba)H_a(X)$.
Proof– $log_b p = log_b a\ log_a p$.⊡

The second property of entropy enables us to change the base of the logarithm in the definition. Entropy can be changed from one base to another by multiplying by the appropriate factor.

And let X be a random variable with cumulative distribution function $F(x) = Pr(X \leq x)$. If $F(x)$ is continuous, the random variable is said to be continuous. Let $f(x) = F'(x)$ when the derivative is defined. If $\int_{-\infty}^{\infty} f(x) = 1$, $f(x)$ is called the probability density function for X. The set where $f(x) > 0$ is called the support set of X. For a continuous random variable X with a probability density function $f(x)$ on S, its entropy is defined as

$$h(X) = -\int_S f(x)logf(x)dx, \tag{2.3}$$

where S is the support set of the random variable. As in the discrete case, the differential entropy depends only on the probability density of the random variable, and therefore the differential entropy is sometimes written as $h(f)$ rather than $h(X)$.

Other information measures

In this section we introduce three related concepts: relative entropy, mutual information, and joint differential entropy. We now extend the definition of these two familiar quantities, $D(f||g)$ and $I(X;Y)$, to probability densities.

2.2.2 Relative entropy

The relative entropy is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(f||g)$ is a measure of the inefficiency of assuming that the distribution is g when

the true distribution is f. The relative entropy (or Kullback–Leibler distance) $D(f||g)$ between two densities f and g is defined by

$$D(f||g) = -\int_S f(x) log \frac{f(x)}{g(x)} dx, \tag{2.4}$$

Note that $D(f||g)$ is finite only if the support set of f is contained in the support set of g. [Motivated by continuity, we set $0 log \frac{0}{0} = 0$.] In the above definition, we use the convention that $0 log \frac{0}{0} = 0$ and the convention (based on continuity arguments) that $0 log \frac{0}{g} = 0$ and $f log \frac{f}{0} = \infty$. Thus, if there is any symbol x on S such that $f(x) > 0$ and $g(x) = 0$, then $D(f||g) = \infty$. The relative entropy is always non-negative and is zero if and only if $f = g$.

2.2.3 Mutual information

The mutual information $I(X;Y)$ between two random variables with joint density $f(x,y)$ is defined as

$$I(X;Y) = \int f(x,y) log \frac{f(x,y)}{f(x)f(y)} dxdy. \tag{2.5}$$

From the definition it is clear that

$$\begin{aligned} I(X;Y) &= h(X) - h(X|Y) \\ &= h(Y) - h(Y|X) \\ &= h(X) + h(Y) - h(X,Y) \end{aligned} \tag{2.6}$$

and

$$I(X;Y) = D(f(x,y)||f(x)f(y)). \tag{2.7}$$

The properties of $D(f||g)$ and $I(X;Y)$ are the same as in the discrete case.

2.2.4 Joint Differential Entropy

The differential entropy of a set $X_1, X_2, ..., X_n$ of random variables with density $f(x_1, x_2, ..., x_n)$ is defined as

$$h(X_1, X_2, ..., X_n) = -\int_S f(x^n) log f(x^n) dx^n, \tag{2.8}$$

2.3 Principle of the Maximum Entropy

The principle of maximum entropy, as a method of statistical inference, is due to Jaynes (Jaynes, 1957a). His idea is that this principle leads to the selection of a probability density function that is consistent with our knowledge and introduces no unwarranted information. Any probability density function satisfying the constraints that has smaller entropy will contain more information (less uncertainty), and thus

says something stronger than what we are assuming. In fact, to the contrary, the principle of maximum entropy guides us to the best probability distribution that reflects our current knowledge and it tells us what to do if experimental data does not agree with predictions coming from our chosen distribution (Conrad, 2004). The Principle of Maximum Entropy is based on the premise that when estimating the probability distribution, you should select that distribution which leaves you the largest remaining uncertainty (i.e., the maximum entropy) consistent with your constraints. That way you have not introduced any additional assumptions or biases into your calculations (Thomas and Cover, 2006).

Finally, the maximum entropy principle states that, for a given amount of information, the probability distribution which best describes our knowledge is the one that maximizes the Shannon's entropy subjected to the given evidence as constraints.

2.3.1 Maximum Entropy Method

The entropy has its maximum value when all probabilities are equal (we assume the number of possible states is finite), and the resulting value for entropy is the logarithm of the number of states, with a possible scale factor ((Thomas and Cover, 2006)). If we have no additional information about the system, then such a result seems reasonable. However, if we have additional information then we ought to be able to find a probability distribution that is better in the sense that it has less uncertainty.

The maximum entropy method (MEM) is determination of a distribution that maximizes the information entropy. The MEM is a great tool for reconstructing a probability distribution given a finite number of moment constraints. This method is typically known as the moments problem. The concept of maximum entropy method was first proposed by Jaynes (Jaynes, 1957a). It is a graceful means of reconstructing a density distribution given a finite number of moment constraints from incomplete data sets.

A general approach for the maximum entropy problem is to maximize Shannon's entropy (Shannon, 1948b) subject to the moment constraints

$$E(\phi_k(X)) = \int \phi_k(x) f(x) dx = \mu_k, \tag{2.9}$$

or

$$\begin{cases} f(x) \geq 0, \\ \int_S f(x) dx = 1, \\ \int_S f(x) \phi_k(x) dx = \mu_k, \quad 1 \leq k \leq m. \end{cases} \tag{2.10}$$

where $\mu_0 = 1$, $\phi_0(x) = 1$, $\phi_k(x)$, $k = 0, ..., N$ are $N + 1$ known functions, and $\mu_k, k = 0, ..., N$ are the given expectation data. $\phi_k(x), k = 0, ..., N$ can be in any functional form such as x^n,$\log(x)$, $x \log(x)$, or trigonometric or geometric functions. The main vehicle to determine the required known function is the relationship of the maximum entropy distribution with the Exponential family (Casella and Berger, 2002; Ebrahimi et al., 2008; Pougaza and Djafari, 2011) using the moment constraints $\mu_1, ..., \mu_m$, which typically can be obtained numerically from data using Taylor's theorem (Casella and Berger, 2002).

Using the method of Lagrange multipliers method, where the objective function is Shannon's entropy Eq.(2.3), $J(f)$ is as follows:

$$J(f) = -\int f(x)logf(x)dx + \lambda_0 \int f(x)dx + \sum_{k=1}^{N} \lambda_k \int f(x)\phi_k(x)dx. \tag{2.11}$$

For obtaining $f(x)$, we differentiate J subject to $f(x)$:

$$\frac{\partial J(f)}{\partial f(x)} = -logf(x) - 1 + \lambda_0 + \sum_{k=1}^{N} \lambda_k \phi_k(x). \tag{2.12}$$

Setting Eq.(2.12) equal to zero, the general form of the maximizing density is obtained as follows:

$$f(x) = e^{-\sum_{k=0}^{N} \lambda_k \phi_k(x)}, x \in S, \tag{2.13}$$

where λ_k should be chosen such that $f(x)$ in Eq.(2.13) satisfies the known moment constraints in Eq.(2.10). In general, there are infinitely many continuous distributions whose constraints match these known constraints. Additional constraints are then required to guide the process of finding a continuous distribution that fits the known constraints. The maximum entropy approach suggests only the form of the density that maximizes the entropy. Thomas and Cover (2006) proved that the distribution in Eq.(2.13) has the maximum entropy.

Definition–(Information inequality) If g satisfies Eq.(2.9) and if $f^\star$ is of the form Eq.(2.13), then $0 \leq D(g||f^\star) = -h(g) + h(f^\star)$. Thus $h(g) \leq h(f^\star)$ for all g satisfying the constraints. We prove this in the following theorem.

Theorem–(Maximum entropy distribution)
Let $f^\star(x) = f_\lambda(x) = e^{\lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x)}$, $x \in S$, where $\lambda_0, ..., \lambda_m$ are chosen so that $f^\star$ satisfies Eq.(2.9). Then $f^\star$ uniquely maximizes $h(f)$ over all probability densities f satisfying constraints Eq.(2.9).

Definition–If a function $g(x)$ has derivatives of order r, that is $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$exists, then for any constant a, the Taylor's polynomial of order r about a is

$$T^r(x) = \sum_{i=0}^{r} \frac{g^i(a)}{i!}(x-a)^i. \tag{2.14}$$

Theorem–(Taylor's Theorem) If $g^{(r)}(a) = \frac{d^r}{dx^r} g(x)|_{x=a}$ exists, then

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0. \tag{2.15}$$

Taylor's major theorem is that the reminder form the approximation, $g(x) - T_r(x)$, always tends to 0 faster than the highest order explicit term Casella and Berger (2002).

To solve the current problem in Eq.(2.13), a more general procedure is clearly needed, and this is provided by the use of Lagrange Multipliers. The technique of Lagrange Multipliers is named after the French mathematician, Joseph-Louis Lagrange (1736 - 1813). Instead of using the constraint equations to reduce the number of unknowns, we increase the number of unknowns. The Lagrange multipliers $\lambda = [\lambda_0, ..., \lambda_N]$ should be calculated to find the class of maximum entropy distributions. To obtain the $N + 1$ Lagrange multipliers, the following set of $N + 1$ nonlinear equations should be solved ($1 \leq k \leq m$):

Here, we use the standard Newton's method, in which the Lagrange multipliers λ are obtained by solving the nonlinear equations. Newton's method relies on a first-order Taylor approximation around trial values of lambda and solving the resulting linear system iteratively (Pougaza and Djafari, 2011; Yari and Farsani, 2015).

$$G_k(\lambda) = \int \phi_k(x) e^{-\sum_k \lambda_k \phi_k(x)} dx = \mu_k. \tag{2.16}$$

Eq.(2.16) can be written in the following matrix form

$$\begin{pmatrix} \mu_0 - G_0 \\ \mu_1 - G_1 \\ \vdots \\ \mu_N - G_N \end{pmatrix} = \begin{pmatrix} \lambda_0 - \lambda^0 \\ \lambda_1 - \lambda^1 \\ \vdots \\ \lambda_N - \lambda^N \end{pmatrix} \begin{pmatrix} \frac{\partial G_0}{\partial \lambda_0} & \frac{\partial G_0}{\partial \lambda_1} & \cdots & \frac{\partial G_0}{\partial \lambda_N} \\ \frac{\partial G_1}{\partial \lambda_0} & \frac{\partial G_1}{\partial \lambda_1} & \cdots & \frac{\partial G_1}{\partial \lambda_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_N}{\partial \lambda_0} & \frac{\partial G_N}{\partial \lambda_1} & \cdots & \frac{\partial G_N}{\partial \lambda_N} \end{pmatrix} \tag{2.17}$$

Since an indisputable analytical solution does not exist for the nonlinear system of Eq.(2.16) in case $N \geq 2$, one must use a nonlinear optimization technique to determine the Lagrange multipliers. One way to solve the maximum entropy problem is to transform the constrained optimization problem into an unconstrained optimization problem via the dual approach (Golan et al., 1997). However, we suggest the standard Newton's method.

2.3.2 Newton's Method

The solution of the standard ME problem is given by Eq.(2.16) in which the Lagrange multipliers, λs, are obtained by solving the nonlinear system of Eq.(2.16). These equations are solved by the standard Newton's method which is consisted of expanding $G_k(\lambda)$ in Taylor's series around trial values of λ_ks, dropping the quadratic and higher-order terms, and solving the resulting linear system, iteratively. We give the details of our numerical method that we implemented. After expanding $G_k(\lambda)$ in Eq.(2.16) by Taylor's series around the trial λ^0, the resulting linear equations are given by;

$$G_k(\lambda) \cong G_k(\lambda^0) + (\lambda - \lambda^0)^t [grad\, G_k(\lambda)]_{(\lambda=\lambda_0)} = \mu_k, \quad k = 0, ..., N. \tag{2.18}$$

Consider the vectors $\delta = \lambda - \lambda^0$ and $\nu = [\mu_0 - G_0(\lambda^0), ..., \mu_N - G_N(\lambda^0)]^t$ and the matrix

G by

$$G = (g_{nk}) = (\frac{\partial G_k(\lambda)}{\partial \lambda_n})_{(\lambda=\lambda_0)}, \quad n, k = 0, ..., N. \tag{2.19}$$

Therefore, Eq.(2.18) becomes

$$\nu = \delta G \tag{2.20}$$

The linear algebraic system in Eq.(2.20) is solved for δ from which we derive$\lambda = \lambda^0 + \delta$, which becomes our initial vector $\lambda^0$and the iteration continues until δbecomes appropriately small. Note that the matrix G is the symmetric one and we have

$$g_{nk} = g_{kn} = -\int \phi_k(x)\phi_n(x) \exp[-\sum_{k=0}^{N} \lambda_k \phi_k(x)]dx, \quad n, k = 0, ..., N. \tag{2.21}$$

Now, the unknown Lagrange multipliers λ_k, $k = 0, ..., N$, are determined by solving Eq.(2.21). A computational procedure is introduced in the next subsection for determining λ_k.

2.4 Joint Maximum Entropy Distribution

Entropy maximization of a joint distribution subject to given marginals has been studied in statistical and probabilistic literature since the 1930s (Cramér and Wold, 1936). The condition for existence of the solution has also been known (Strassen et al., 1965). This problem was also considered in (Kullback, 1968) and (Csiszár, 1997). Two years before Sklar's theorem was published, Edwin Jaynes proposed, in two seminal papers (Jaynes, 1957a,b), the Principle of Maximum Entropy which defines probability distributions given only partial information. Principle of Maximum Entropy has been used in many areas and originally when the partial information is in the form of knowledge of some geometric or harmonic moments (e.g.(DJAFARI, 1994; Mohammad-Djafari, 1991)). The case where the entropy considered is the Shannon's entropy on a measurable space was discussed more rigorously in (Borwein et al., 1994), and this idea was later used in Meeuwissen and Bedford (1997), where the authors derive the joint distribution with given uniform marginals on $I = [0, 1]$ and given correlation. Here the partial information is the knowledge of the marginal distributions. The main result is that we can determine a multivariate distribution with given marginals and which maximizes an entropy (Pougaza and Djafari, 2011).

2.4.1 Inverse Ill-posed Problem

The terms "inverse problems" and "ill-posed problems" have been steadily and surely gaining popularity in modern science since the middle of the 20th century. A little more than fifty years of studying problems of this kind have shown that a great number of problems from various branches of classical mathematics (computational algebra, differential and integral equations, partial differential equations, functional analysis) can be classified as inverse or ill-posed, and they are among the most complicated ones

(since they are unstable and usually nonlinear). At the same time, inverse and ill-posed problems began to be studied and applied systematically in physics, geophysics, medicine, astronomy, and all other areas of knowledge where mathematical methods are used. The reason is that solutions to inverse problems describe important properties of media under study, such as density and velocity of wave propagation, elasticity parameters, conductivity, dielectric permittivity and magnetic permeability, and properties and location of inhomogeneity in inaccessible areas, etc (Kabanikhin, 2008).

What are inverse and ill-posed problems? While there is no universal formal definition for inverse problems, an "ill-posed problem" is a problem that either has no solutions in the desired class, or has many (two or more) solutions, or the solution procedure is unstable (i.e., arbitrarily small errors in the measurement data may lead to indefinitely large errors in the solutions). Most difficulties in solving ill-posed problems are caused by the solution instability. Therefore, the term "ill-posed problems" is often used for unstable problems (Kabanikhin, 2008).

In 2D case, interpreting the joint probability density function $f(x, y)$ as an image and its marginal probability densities $f_1(x)$ and $f_2(y)$ as horizontal and vertical line integrals:

$$f_1(x) = \int f(x, y)dy \quad and \quad f_2(y) = \int f(x, y)dx \tag{2.22}$$

we see that the problem of determining $f(x, y)$ from $f_1(x)$ and $f_2(y)$ is an ill-posed (inverse) problem (Hadamard, 1902; Tarantola, 2005). It is a well known fact that while a distribution has a unique set of marginals, the converse is not necessarily true. That is, many distributions may share a common subset of marginals. In general, it is not possible to uniquely reconstruct a distribution from its marginals.

Determining $f(x, y)$ from $f_1(x)$ and $f_2(y)$ is an ill-posed undetermined inverse problem and in statistics, the notion of copula is exactly introduced to characterize all the possible solutions to the problem of reconstructing a bivariate density from its marginals (Pougaza and Djafari, 2011).

2.4.2 Copula

Copulas are receiving much interest from various research fields such as risk management (Spaces, 1983), networked systems (Marshall and Olkin, 1967), fuzzy systems (Genest and MacKay, 1986), (Joe, 1993), signal processing (Jaynes, 1957b) and modeling of time series (Jaynes, 1957a), (Mohammad-Djafari, 1991). Copulas provide an easy way to construct joint distribution functions splitting the task in two ways: 1) estimating the margins of each random variable separately, for which there are well known methods and plenty of software aids, and 2) estimating the dependence structure, paying no attention to the margins. This separation allows testing different copulas until the best fit is found. On the other hand, when the dependence structure is non-linear, as often is the case in real situations, using the linear correlation as a measure of dependence is an erroneous assumption. The Archimedean family of copulas has the nice property of being directly linked with non-linear degrees of association, such as Kendall's τ or Spearman's ρ.

Copulas have proven useful for modeling the dependence structure between variables in the presence of partial information: the knowledge of marginal distributions. In other word, a copula is a multivariate probability distribution which describes the dependence between random variables. Here, we are interested in finding the bivariate distribution, knowing only its marginal distributions. This is an ill-posed inverse problem (Tarantola, 2005) as it does not have a unique solution.One possible way to select a unique solution to this problem is to choose an appropriate copula and use Sklar's theorem (Rüschendorf, 2013) according to which there exists a copula which relates the marginal distributions yielding to the joint distribution. The problem then becomes the choice of a copula.

Theorem–(Sklar's Theorem) Let F be a joint distribution function with marginal distribution F_1 and F_2. There exists a copula C such that, for all $x, y \in (-\infty, \infty)$,

$$F(x, y) = C(F_1(x), F_2(y)), \tag{2.23}$$

If F_1 and F_2 are continuous, the copula C is unique; otherwise, C is uniquely determined on (Range of F_1) $\times$ (Range of F_2). On the other hand, if C is a copula and F_1 and F_2 are univariate distribution functions, F is a joint distribution function with marginal distributions F_1 and F_2, (Balakrishnan and Lai, 2009). $\boxdot$

According to Sklar's theorem, there exists a copula which relates the marginal distributions yielding the joint distribution. The remaining problem is choice of a copula. Note that there are many other ways to derive families of continuous multivariate distributions with given univariate marginals (Genest and MacKay, 1986; Genest and Rivest, 1993; Pougaza and Djafari, 2011).

By $F(x, y)$ we denote a continuous bivariate cumulative distribution function (CDF), and $f(x, y)$ its bivariate probability density function (PDF). Let $F_1(x)$, $F_2(y)$ be the marginal CDFs and $f_1(x)$, $f_2(y)$ their respective PDFs. A bivariate copula C is a function from $[0, 1]^2$ to $[0, 1]$ with the following properties:

- $\forall u, v \in [0, 1], C(u, 0) = 0 = C(0, v)$,
- $\forall u, v \in [0, 1], C(u, 1) = u \; and \; C(1, v) = v$,
- $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all $u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2, v_1 \leq v_2$.

One can construct copulas C from joint distribution functions by

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \tag{2.24}$$

where the quantile function is $F_i^{-1}(t) = inf\{u : F_i(u) \geq t\}$ (Pougaza and Djafari, 2011; Romeo et al., 2006). For any multivariate absolutely continuous distribution, with CDF F and marginal CDFs $F_i : \mathbb{R} \mapsto [0, 1]$, the copula C is such a distribution function on $[0; 1]^p$ with uniform one-dimensional marginal distributions such that

$$F(x_1, ..., x_p) = C(F_1(x_1), ..., F_p(x_p)), \tag{2.25}$$

which is a joint probability distribution for the vector-valued random variables $X = (X_1, ...X_m)$ with the joint copula density $c : [0,1]^p \mapsto [0,\infty)$ defined by

$$c(F_1(x_1), ..., F_p(x_p)) = \frac{\partial^p C(F_1(x_1), ..., F_p(x_p))}{\partial F_1 ... \partial F_p}. \tag{2.26}$$

The joint density $f : \mathbb{R}^p \mapsto [0,1]$ is defined almost everywhere and it can be expressed as (Balakrishnan and Lai, 2009)

$$f(x) = c(F_1(x_1), ..., F_p(x_p)) \prod_{i=1}^{p} f_i(x_i). \tag{2.27}$$

Fitting a copula to a given independent identically distributed (*i.i.d.*) data set of n observations (x_i, y_i) is a field of intense research (Burg, 1975; Havrda and Charvát, 1967; Rényi et al., 1961). The general approach takes advantage of the separation of margins and copula provided by Sklar's theorem. Thus, in a first step the margin functions are estimated either with parametric or non parametric methods. The result are parametric or empirical CDF margins respectively. Then in a second step, the estimation of the copula's parameter a is carried out via Maximum Likelihood (ML) (Strassen et al., 1965).

Archimedean copulas have the advantage of being characterized by the single-valued generator ϕ, regardless of how many random variables are coupled, and one can then skip the estimation of the margins. Thus, Genest and Rivest (Rényi et al., 1961) propose a parametric approach: assuming that the dependence structure is captured by an Archimedean copula, the goal is to choose the generator and compute its parameter.

Property– Any copula $C(u, v)$, satisfies the inequality

$$W(u, v) \leq C(u, v) \leq M(u, v), \tag{2.28}$$

where the **Fréchet–Hoeffding upper bound copula** $M(u, v)$ (or comonotonicity copula) is:

$$M(u, v) = min(u, v), \tag{2.29}$$

and the **Fréchet–Hoeffding lower bound** $W(u, v)$ (or countermonotonicity copula) is:

$$W(u, v) = max\{u + v - 1, 0\} \ (u, v) \in [0, 1]^2. \tag{2.30}$$

Generating copulas by the inversion method: A straightforward method is based directly on Sklar's theorem. Given $F(x, y)$ the joint cdf of two random variables X, Y and $F_1(x)$ and $F_2(y)$ their marginal cdf's, all assumed to be continuous. The corresponding copula can be constructed by using the unique inverse transformations (Quantile transform) $x = F_1^{-1}(u), y = F_2^{-1}(v)$, and the Eq. (2.24) where u, v are uniform on $[0, 1]$.

2.4.3 Archimedean Copula Model

The Archimedean copulas (Nelsen, 2007) form an important class of copulas which generalize the usual copulas.

Theorem– Let ϕ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\phi(1)=0$, and let $\phi^{[-1]}$ be the pseudo-inverse1 of ϕ.

$$\phi^{[-1]}(t)=\begin{cases}\phi^{[-1]}(t),\ 0\leq t\leq\phi(0)\\ 0,\ \phi(0)\leq t\leq\infty\end{cases} \tag{2.31}$$

Let C be the function from $[0,1]^2$ to $[0,1]$ given by

$$W(u,v)=max\{u+v-1,0\}\ (u,v)\in[0,1]^2. \tag{2.32}$$

Then C is a copula if and only if ϕ is convex.

Archimedean copulas are in the form Eq.(2.32) and the function ϕ is called the generator of the copula. ϕ is a strict generator if $\phi(0)=\infty$, then $\phi^{[-1]}=\phi^{-1}$ and

$$C(u,v)=\phi_{-1}(\phi(u)+\phi(v)). \tag{2.33}$$

Property– Any Archimedean copula C satisfies the following algebraic properties:

- $C(u,v)=C(v,u)$ meaning that C is symmetric;
- $C(C(u,v),w)=C(u,C(v,w))$;
- If $a>0$, then $a\phi$ is again a generator of C.

There are many families of Archimedean copulas constructed from different generators ϕ_α with a suitable parameter α. For example $\phi_\alpha(t)=\frac{1}{\alpha}(t^{-\alpha}-1)$ and $\phi_\alpha(t)=ln(1-\alpha ln(t))$ yield successively to Clayton copula $C_\alpha(u,v)=[max(u^{-\alpha}+v^{-\alpha}-1,0)]^{\frac{1}{\alpha}}$ and Gumbel-Hougaard copula $C_\alpha(u,v)=uv\times exp(-\alpha lnu\times lnv)$.

2.4.4 Gaussian Copula Model

In our proposed method, we use a Gaussian copula. A Gaussian copula is a distribution over the unit cube $[0,1]^d$ constructed from a multivariate Gaussian distribution over $\mathbb{R}$ by using the probability integral transform. For a given correlation matrix $\mathbf{R}\in[-1,1]^{d\times d}$, the Gaussian copula with parameter matrix $\mathbf{R}$ can be written as

$$C_{\mathbf{R}}^{\text{Gauss}}(u)=\Phi_R\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right), \tag{2.34}$$

where Φ^{-1} is the inverse CDF of a standard Gaussian and $\Phi_{\mathbf{R}}$ is the joint CDF of a multivariate Gaussian distribution with mean vector zero and covariance matrix equal to the correlation matrix $\mathbf{R}$. While there is no simple analytical formula for the copula function, $C_{\mathbf{R}}^{\text{Gauss}}(u)$, it can be upper or lower bounded, and approximated using numerical integration (Botev, 2017). The density of the proposed copula can be written as:

$$c_{\mathbf{R}}^{\text{Gauss}}(u)=\frac{1}{\sqrt{\det\mathbf{R}}}\exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots\\ \Phi^{-1}(u_d)\end{pmatrix}^T\cdot\left(\mathbf{R}^{-1}-\mathbf{I}\right)\cdot\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots\\ \Phi^{-1}(u_d)\end{pmatrix}\right), \tag{2.35}$$

where $\mathbf{I}$ is the identity matrix (Arbenz, 2013).

2.4.5 Maximum Entropy Copula

In order to find the bivariate maximum entropy PDF $f(x, y)$, the marginal distributions become the constraints:

$$\begin{cases} \phi_0(x,y) : \int\int f(x,y)dxdy = 1, \\ \phi_1(x,y) : \int f(x,y)dy = f_1(x),\ \forall x, \\ \phi_2(x,y) : \int f(x,y)dx = f_2(y),\ \forall y. \end{cases} \tag{2.36}$$

Hence, the goal is to find the bivariate density $f(x, y)$ compatible with available information in the sense of the maximum entropy principle. Among all possible $f(x, y)$ satisfying the constraints Eq.(2.36), we select the one which optimizes the entropy $h(X, Y)$:

$$h(X,Y) = -\int\int f(x,y)\log f(x,y)dxdy \tag{2.37}$$

via

$$\hat{f} := maximize\ h(X,Y)\ subject\ to\ \phi_k s\ in\ Eq.(2.36) \tag{2.38}$$

Because the constraints are linear, the choice of a concave objective function $h(X, Y)$ guarantees the existence of a unique solution to the problem. Many entropy functions can serve as concave objective functions, but we focus on the Shannon's entropy (Shannon, 1948b)

$$f(x,y) = exp(-\lambda_1\phi_1(x,y) - \lambda_2\phi_2(x,y) - \lambda_0\phi_0(x,y)) \tag{2.39}$$

where λ_1, λ_2 and λ_0 are obtained by replacing these expressions in the constraints Eq.(2.36) and solving the resulting system of equations. For Shannon's entropy, the constraints can be solved analytically and the joint distribution becomes

$$f(x,y) = f_1(x)f_2(y). \tag{2.40}$$

With the bivariate density obtained from the maximum entropy principle, we can immediately find the corresponding bivariate copula. For the case of Shannon's entropy we have:

$$\begin{aligned} F(x,y) &= \int_0^x\int_0^y f(s,t)dsdt \\ &= \int_0^x\int_0^y f_1(s)f_2(t)dsdt \\ &= \int_0^x f_1(s)ds\int_0^y f_2(t)dt \end{aligned} \tag{2.41}$$

The CDF becomes

$$F(x,y) = F_1(x)F_2(y), \tag{2.42}$$

and the copula is

$$C(u,v) = F(F_1^{-1}(u), F_2^{-1}(v)) = uv. \tag{2.43}$$

In this case, the maximum entropy copula obtained from the Shannon' entropy is the well-known independent copula, which describes independence between two random variables (Pougaza and Djafari, 2011).

In general, the properties of Pearson's product moment correlation coefficient in bivariate Normal samples (bi–normal model) is well known thanks to the creative work of Fisher (1922). However, sometimes it might not be applicable when the following scenarios happen:

- The data are incomplete, that is, only ordinal information is available. This situation is not uncommon in the area of social sciences, such as psychology and education (Kendall, 1948);
- The underlying data are complete (cardinal) and follow a bivariate Normal distribution, but is attenuated more or less by some monotone non-linearity in the transfer characteristics of sensors (Tumanski, 2006);
- The data are complete and the majority follow a bivariate Normal distribution, but there exists a tiny fraction of outliers (impulsive noise) (Stein, 1995; Xu et al., 2010).

Under these circumstances, it would be more suitable to employ the most popular non-parametric coefficient, Kendall's τ, which is 1) dependent only on ranks, 2) invariant under increasing monotone transformations (Kendall, 1948), and 3) robust against impulsive noise (Shevlyakov and Vilchevski, 2002).

2.4.6 Kendall's τ

Let (x_i, y_i) and (x_j, y_j) be two observations from (X, Y), a pair of continuous random variables. The two pairs (x_i, y_i) and (x_j, y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) \geq 0$ and discordant if $(x_i - x_j)(y_i - y_j) \leq 0$. Kendall's τ is defined as the probability of concordance minus the probability of discordance,

$$\tau = P[(X - X')(Y - Y') \geq 0] - P[(X - X')(Y - Y') \leq 0] \tag{2.44}$$

The definition above is equivalent to

$$\tau = cov[sgn(X' - X), sgn(Y' - Y)]. \tag{2.45}$$

τ may also be defined as

$$\tau = 4 \int \int C(x, y)c(x, y)dxdy - 1. \tag{2.46}$$

The sample version of Kendall's τ is defined as

$$t = \frac{c - d}{c + d} = \frac{c - d}{n} \tag{2.47}$$

where c denotes the number of concordant pairs and d the number of discordant pairs from a sample of n observations from (X, Y). $\hat{\tau}$ is an unbiased estimator of τ. Just as H can be expressed as a function of copula C, Kendall's τ can be expressed in terms of the copula [see, for example, Nelsen (2007)] as

$$\tau = 4 \int_0^1 \int_0^1 C(u, v)c(u, v)dudv - 1 = 4E(C(U, V)) - 1. \tag{2.48}$$

2.4.7 Copula for Non-Gaussian Model

Radon's Transformation

Copula is useful for constructing joint distributions, particularly with non-Gaussian random variables (Joe, 1997). In 2D case, interpreting the joint probability density function $f(x, y)$ as an image and its marginal probability densities $f_1(x)$ and $f_2(y)$ as horizontal and vertical line integrals:

$$f_1(x) = \int f(x, y)dy \quad and \quad f_2(y) = \int f(x, y)dx$$

we see that the problem of determining $f(x, y)$ from $f_1(x)$ and $f_2(y)$ is an ill-posed (inverse) problem (Hadamard, 1902; Idier, 2013; Tarantola, 2005). It is a well known fact that while a distribution has a unique set of marginals, the converse is not necessarily true. That is, many distributions may share a common subset of marginals. In general, it is not possible to uniquely reconstruct a distribution from its marginals.

The case of maximum entropy method in image processing has been studied by Pougaza et al. (2010). He first proposes a brief description of the maximum entropy method priors to solve the linear system of equations which is obtained after the discretization of the integral equations, which arises in various tomographic image restoration and reconstruction problems. Then he discusses the main problem which is to choose an a priori probability law for the image and to determine their parameters from the data and he suggests then a method to estimate simultaneously the parameters of the maximum a priori probability density function (pdf) and the pixel' values of the image and give some simulated results which compare this method with some classical ones. In the second case, he has been noted that an important problem in statistics is construction of a joint probability distribution from its marginal. He linked this issue to an important problem in Computed tomography (CT) as a reconstruction an image from its projections. For determining $f(x, y)$ from $f_1(x)$ and $f_2(y)$, he proposed the using copula by maximum entropy method as the solution to the problem of reconstructing a bivariate density from its marginals (Pougaza and Djafari, 2011). In 1917 Johann Radon (Radon, 1917) introduced the radon transform (RT) which was later used in CT. Indeed if we denote by $f(x, y)$, the spatial distribution of the material density in a section of the body, a very simple model that relates a line of radiography image $q(r, \theta)$ at direction θ to $f(x, y)$ is given by Radon transform:

$$q(r, \theta) = \int_{L_{r,\theta}} p(x, y)dl = \int \int_{R^2} p(x, y)\delta(r - xcos\theta - ysin\theta)dxdy. \tag{2.49}$$

where $L_{r,\theta} = (x, y) : r = xcos\theta + ysin\theta$ and δ is the Dirac's delta function. In 2D, the mathematical problem of tomography is to determine the bivariate function $f(x, y)$ from its line integrals $q(r, \theta)$. Radon has shown that this problem has a unique solution if we know $q(r, \theta)$ for all $\theta \in [0, \pi]$ and all $r \in R$, then $f(x, y)$ can be computed by the inverse radon transform:

$$f(x, y) = (-\frac{1}{2\pi}) \int_0^{\pi} \int_{-\infty}^{+\infty} [\frac{\frac{\partial q(r,\theta)}{\partial r}}{r - xcos\theta - ysin\theta}]dr.d\theta \tag{2.50}$$

2.5 Image Processing

Image analysis is playing a very essential role in numerous research areas in the fields of science and technology, ranging from medical imaging to the computer science of automatic vision (Petrovici et al., 2018; Sparavigna, 2019). Being involved in several applications, which are mainly based on a constant innovation of technologies, image analysis always requires different approaches and new algorithms, including continual upgrades and improvements of existing methods. Accordingly, the range of the analyses in which it is engaged can be considered as wide as the prospective future technologies. A challenge of image analysis is obtaining meaningful information by extracting specific features from images contained in large databases or during real-time acquisition. These tasks demand highly sophisticated numerical and analytical methods.

In the other word, many imaging systems are faced with the problem of estimating a true image from a degraded data set. In such systems, the image degradation is translated into a convolution with a Point Spread Function (PSF) and addition of noise. Often, the image recovery by inverse filtering is not possible because the PSF matrix is ill-conditioned (Denisova, 2019; Sparavigna, 2019). There is a huge amount of literature describing methods, which concerns quality improvement of the image information content through restoration techniques Jia et al. (2019); Liu et al. (2019); Petrovici et al. (2018); Sparavigna (2019). In most of the contributions, entropy plays a pivotal role (Denisova, 2019; Sparavigna, 2019).

Image reconstruction belongs to the class of ill-posed inverse problems of mathematical physics (Hadamard, 1932). The method of solution for this class of problems was developed in 1963 by A.Tikhonov and called regularization (Tihonov, 1963). Regularization introduces a priori information regarding the problem to obtain well-behaved inverse. The introduction of a priori information entails the necessity of choosing an unknown parameter, which is called the regularization parameter. In Tikhonov's approach, this parameter plays a role of a weight factor, which controls the competition between a priori information and the measured data. Initially, the method was developed in the form of 'global regularization', in which a single parameter controls the solution. However, it was found that global regularization often provides too smoothed solutions, even in the early practical applications (Arsenin and Timonov, 1983; Tikhonov et al., 1984).

In 1967, the physicist V. Turchin suggested using the Bayesian method of Maximum a Posteriori (MAP) for solving inverse ill-posed problems with stochastic data, naming this approach 'statistical regularization' (Turchin et al., 1971; Turchin, 1967). The statistical regularization method introduces a priori information in the probabilistic form. Later, in the 70–80s the Bayesian method for solving image restoration and image reconstruction problems had become popular. The main difficulty of the Bayesian approach is the determination of a priori probability density function. Two forms of prior probability are used for solving the problems of restoring and reconstruction of images: entropy concept and Gibbs probability distribution. Jaynes (Jaynes, 1957a,b,

1968) suggested prior information, based on the entropy concept, for solving problems with limited, but noise-free data. This approach was called the maximum entropy method.

In 1972, the ME approach was applied by Frieden to solving the image restoration problem (Frieden, 1972). After a few years, the ME approach was successfully applied in the restoration of radio astronomy images in the paper by Gull and Daniell (Gull and Daniell, 1978). In (Frieden, 1972; Gull and Daniell, 1978), the image restoration problem was solved as a constrained optimization problem. J. Skilling developed the approach based on the Bayesian method Maximum a Posteriori with entropy-based prior probability (Skilling, 1988). In 1979, Minerbo used the ME approach for the tomography problems (Minerbo, 1979). The tomographic task was solved as constrained optimization problem. The probabilistic approach entropy-based prior probability for solving tomographic problems was developed and applied to plasma tomography by W. von der Linden (von der Linden, 1995). In (Denisova, 2004b), the relation between entropy-based prior probability and ME was studied. An improvement in reconstructed image quality by the MAP-ENT algorithm over the ME was demonstrated. In (Denisova, 2004a), the entropy-based prior probability approach was developed and applied for nuclear medicine.

Besag (Besag, 1974) and D. Geman and S. Geman (Geman and Geman, 1984) theoretically justified another form of a priori probability known as 'the Gibbs prior'. The approach that was based on the Bayesian method Maximum a Posteriori with Gibbs a priori probability (MAP-GIBBS) is widely studied for applications in nuclear medicine. Geman and McClare first applied this approach to nuclear medicine in 1985 (Ganan and McClure, 1985). Denisova (2019) provided a solution for this problem by combination of the maximum a posteriori (MAP) Bayesian approach with maximum entropy method.

2.5.1 Bayesian approach Maximum a Posteriori (MAP) & Maximum Entropy Method

The Bayesian approach Maximum a Posteriori (MAP) provides a common basis for developing statistical methods for solving ill-posed image reconstruction problems. MAP solutions are dependent on a priori model. Approaches developed in literature are based on prior models that describe the properties of the expected image rather than the properties of the studied object. In this paper, such models have been analyzed and it is shown that they lead to global regularization of the solution. Prior models that are based on the properties of the object under study are developed and conditions for local and global regularization are obtained (Mohammad-Djafari, 1996).

A class of discrete image-reconstruction and restoration problems is addressed. A brief description is given of the maximum a posteriori (MAP) Bayesian approach with maximum entropy priors to solve the linear system of equations which is obtained after the discretization of the integral equations which arises in various tomographic image restoration and reconstruction problems (Djafari and Demoment, 1990).

The main difficulty of the Bayesian approach is the determination of the a priori probability law for the image and determining its parameters from the data. There

are two different classes of imaging problems: restoration of distorted images and tomographic reconstruction of objects. When solving the image restoration problems, it is usually impossible to include a priori information about real objects. A priori information refers to the expected image to be restored. In contrast, when reconstructing tomographic images, one can often propose the physical model of the object to be reconstructed and determine a priori information by using this model. Currently, in tomographic problems, one uses a priori information, which was initially developed to solve the image restoration problems. In this paper, we discuss these approaches to specifying a priori probability in both the image restoration problem and the image reconstruction problem. We consider the three most widely used forms of a priori information: no prior, Gibbs prior, and entropy-based prior (Denisova, 2019; Mohammad-Djafari, 1996; Skilling, 1988; Sparavigna, 2019).

A method simultaneously estimating the parameters of the ME a priori probability density function and the pixel values of the image is proposed, and some simulations which compare this method with some classical ones are given.

2.5.2 Maximum A Posterior Probability Approach

The Bayesian approach Maximum a Posteriori (MAP) prepared a common basis in statistical methods to solve an ill-posed image reconstruction problems. The main thing is that, solutions of MAP are highly dependent on a priori model. The proposed methods developed in literature are based on prior models that describe the properties of the expected image rather than the properties of the studied object. The main difficulty of the Bayesian approach is the determination of the a priori probability law for the image and determining its parameters from the data. In this situation, there are two different classes of imaging problems: restoration of distorted images and tomographic reconstruction of objects. When solving the image restoration problems, it is usually impossible to include a priori information about real objects (Denisova, 2019; Sparavigna, 2019).

In this paper, we discuss these approaches to specifying a priori probability in both the image restoration problem and the image reconstruction problem. We consider the most widely used forms of a priori information which named entropy-based prior, based on the entropy principle, was successfully applied in the fields of X-ray-, radio- and gamma-astronomy, plasma tomography (Denisova, 2019; Mohammad-Djafari, 1996; Skilling, 1988; Sparavigna, 2019).

To estimate the kinetic parameters, considering the general form of Eq.(5.8) as follows (Djafari and Demoment, 1990; Mohammad-Djafari, 1996; Mohammad-Djafari and Demoment, 1990). Estimate the positive vector x (representing the pixel intensities in an object) given a vector of measurement y (representing either a degraded image or the projections of the object) and a linear transformation A relating them by

$$y = Ax + b, \tag{2.51}$$

where b presents the noise measurement, which is supposed to be an uncorrelated normal distribution with zero-mean and additive. We assume we have only approximate information about the noise variance and some global information about the object.

To determine the unknown vector $\vec{x}$, we apply the Bayesian rule. We would like to find the peak of the conditional distribution for x given y, $f(y|x)$, and take the corresponding x-vector as a solution of the reconstruction problem. This is a well-known method in parameter estimation named maximum a posteriori (MAP). This method requires the specification of $f(x)$ and $f(y|x)$. The estimator $\hat{x}$ maximizes the posterior distribution obtained by Bayesian formula (Elfving, 1989):

$$f(x|y) = f(y|x)f(x)/f(y), \tag{2.52}$$

In Eq.(5.14), $f(y)$ is independent of x, $f(y|x)$ is in fact related to the noise probability law and $f(x)$ is a prior law on x. With a linear relation between x and y, and assuming uncorrelated normally distributed errors in the linear relation by knowing the only variance σ^2 of the noise. We are not able to directly determine $f(x)$ and $f(y|x)$. So, we apply maximum entropy method. The idea is if we have not enough information about the random process to determine it a probability law, we choose the maximum entropy principle which satisfies our a priori information. The maximum entropy principle can be used if this knowledge can be stated as some constraints on $f(x)$. In general these only constraints are not sufficient to determine uniquely fx). Then between all probabilities law which satisfied these constraints, we choose the one which has maximum entropy (Djafari and Demoment, 1990; Mohammad-Djafari, 1996; Mohammad-Djafari and Demoment, 1990), see section (5.2.2). It is easy to obtain $f(y|x)$ as a correlated normally distribution via MEM as follows:

$$\begin{aligned} f(y|x) &\approx \exp[-T(x)], \\ T(x) &= [y - Ax]^t[y - Ax]/\sigma^2. \end{aligned}$$

A possibility to choose a priori distribution $f(x)$ is again the MEM with the general model belongs to the Exponential family. The idea of favorite in using this concept is most objective or maximally uncommitted with respect to missing information Elfving (1989). Then the solution of our problem is the mode of the posterior probability distribution as follows:

$$\begin{aligned} \hat{x} &= Arg\max_{x>0} f(x|y) \\ &= Arg\max_{x>0}\{f(y|x)f(x)\} \end{aligned} \tag{2.53}$$

Chapter 3

Maximum Entropy Approach in Dynamic Contrast–Enhanced Magnetic Resonance Imaging

Contributing article Zahra Amini Farsani, Volker J. Schmid (2017). Maximum Entropy Approach in Dynamic Contrast–Enhanced Magnetic Resonance Imaging". *Methods of Information in Medicine*, vol 56, no 6, 2017, pp. 461-468. DOI: 10.3414/ME17-01-0027.

Author contributions Zahra Amini Farsani prepared a first draft of the manuscript. Prof. Schmid added valuable input, particularly on the introduction, method, the application and some parts of the proofs. The asymptotic properties were derived in close collaboration of both authors. All implementations were done by Zahra Amini Farsani.

Chapter 4

Co-localization Analysis in Fluorescence Microscopy via Maximum Entropy Copula

Contributing article Zahra Amini Farsani, Volker J. Schmid (2021). Co-localization Analysis in Fluorescence Microscopy via Maximum Entropy Copula. *The International Journal of Biostatistics*, vol. 17, no. 1, 2021, pp. 165-175. Doi: 10.1515/ijb-2019-0019.

Author contributions Zahra Amini Farsani prepared a first draft of the manuscript. All simulations was done by Prof.Schmid. In general, this work was derived in close collaboration of both authors. All implementations were done by Zahra Amini Farsani.

Chapter 5

Modified Maximum Entropy Method for Estimating the AIF and DCE-MRI Data Analysis

Submitted to Entropy

Author contributions Zahra Amini Farsani and Prof. Schmid conceived of the presented idea. Zahra Amini Farsani developed the theory and performed the computations and verified the analytical methods. Prof. Schmid encouraged Zahra Amini Farsani to investigate different parameter estimation methods and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Modified Maximum Entropy Method for Estimating the AIF and DCE-MRI Data Analysis

Zahra Amini Farsani[1], [2] and Volker J Schmid [2]

[1]Bioimaging Group, Department of Statistics, Ludwig-Maximilians-University of Munich, Ludwig.str 33, 80539, Munich, Germany.

BACKGROUND For kinetic models used in contrast-based medical imaging, the determination of the arterial input function (AIF) is essential for the estimation of physiological parameters of the tissue via solving a nonlinear inverse problem named.
OBJECTIVE In this paper, we estimate the AIF based on the modified maximum entropy method. The effectiveness of several numerical methods to determine kinetic parameters and the AIF is evaluated– in situations where enough information about the AIF is not available. The purpose of this study is to identify one of the most appropriate methods for estimating this function.
MATERIALS AND METHODS The modified algorithm is a combination of the maximum entropy method and an optimization method, which was named the modified maximum entropy method (MMEM). Then, we applied this algorithm in a Bayesian framework to estimate the kinetic parameters via the unique form of the arterial input function. We assessed the efficiency of our algorithm to estimate the kinetic parameters of Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) and AIF with some other parameter-estimation methods and a standard fixed AIF method. A previously analyzed dataset consisting of contrast agent concentrations in tissue and plasma was used.
RESULTS AND CONCLUSIONS To evaluate accuracy, we compared the results obtained from the MMEM with those of the Empirical Method, Maximum Likelihood, the Modified Maximum Likelihood Method, moment matching ("method of moments"), and the least square method. The numerical results indicated the Weibull distribution as an appropriate and robust AIF and also illustrated the power and effectiveness of the proposed method to estimate its parameters.

keywords: Kinetic Model; Modified Maximum entropy method; Arterial input function; Optimization Method;

5.1 Introduction

Determining probability density of a random variable based on observations is a major and old issue in statistics. In recent years, various parametric and non-parametric methods have been introduced for determining the class of different statistical distributions. In general, in the case of continuous random variables, a probability density function (PDF) assigns a probability for the observation of a value falling within a specific given range. Empirically determining a PDF corresponding to N samples of univariate data has been investigated extensively in mathematics, with ubiquitous significance for practical applications. Multiple estimation approaches have been used with success for fitting a random data sample to parameters of a known functional form. Nevertheless, the functional form describing the underlying random process is often unknown. The

maximum entropy method is used today as a major method for estimating and determining the probability density with high accuracy and efficiency and low bias. This method is employed to obtain the unknown density by solving optimization problems and regarded as one of the most efficient methods capable of yielding maximum possible information for unknown density using the limited and known available information (Ebrahimi et al., 2008; Pougaza and Djafari, 2011; Thomas and Cover, 2006).

The maximum entropy framework is a cornerstone of statistical inference, which is employed at a growing rate for constructing models capable of describing and predicting biological systems, particularly complex one, from empirical datasets. Entropy maximization or related concepts have been frequently utilized in the past ten years to analyze large biological datasets in various fields. These fields range from determining macromolecular interactions and structures (Boomsma et al., 2014; Cavalli et al., 2013; Cesari et al., 2018; Cofré et al., 2019; D'haeseleer et al., 2000; Ekeberg et al., 2013; Farré and Emberly, 2018; Hopf et al., 2012; Jaynes, 2003; Jennings et al., 2020; Ozer, 2008; Pitera and Chodera, 2012; Roux and Weare, 2013; Seno et al., 2008; Weigt et al., 2009; Zhang and Wolynes, 2015) to inferring signaling (Dhadialla et al., 2009; Lezon et al., 2006; Locasale and Wolf-Yadlin, 2009; Remacle et al., 2010; Sanguinetti et al., 2019) and regulatory networks (Graeber et al., 2010; Schneidman et al., 2006; Sharan and Karp, 2013) and the coding organization in neural populations (Cocco et al., 2009; Ferrari et al., 2017; Granot-Atedgi et al., 2013; Mora et al., 2010; Nghiem et al., 2018; Ohiorhenuan et al., 2010; Quadeer et al., 2020; Rostami et al., 2017; Roudi et al., 2009; Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Tkačik et al., 2013, 2010; Yeh et al., 2010; Yeo and Burge, 2004) based on DNA sequence analyses (the detection of specific binding sites, for instance) (Fariselli et al., 2020; Fernandez-de Cossio-Diaz and Mulet, 2019; Mora et al., 2010; Santolini et al., 2014; Yeo and Burge, 2004). Here, a concise and basic introduction to entropy maximization and its applicability for deriving models from biological datasets especially in kinetic model and image processing via Dynamic Contrast-Enhanced Magnetic Resonance Imaging data is provided.

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is a fast and noninvasive method for the quantitative analysis of contrast agent (CA) transient in soft tissues. This has made this method an important and well-developed tool to manifest microvasculature and perfusion of blood in a variety of clinical applications. In the last three decades, a host of nonparametric and parametric models and methods have been developed to quantify the CA's perfusion into tissues and estimate perfusion-related parameters (indices) from signal- or concentration-time curves. These indices are widely used in various clinical applications for detecting, characterizing, and therapeutic monitoring of different diseases (Fennessy et al., 2014; Huang et al., 2014; Khalifa et al., 2014; Sobhani et al., 2016; Stoyanova et al., 2012; Usuda et al., 2019).

Quantification of perfusion in DCE-MRI involves the measurement of the time course of gadolinium concentration in the left ventricular blood pool, commonly known as the arterial input function (AIF). In DCE-MRI, a tracer which is made up of the magnetic contrast agent is utilized in situations where the concentration in tissue is imaged over time. The kinetic processes of this contrast agent display the perfusion of blood in this tissue. An important part of the kinetic models used in contrast-based medical imaging is to determine the arterial input function (AIF) which is essential

for the estimation of physiological parameters. Actually, the main difficulty in perfusion acquisition is to accurately measure the AIF, which has been considered in many studies (Larsson and Tofts, 1992).

As we mentioned before, the non-linear kinetic models depend highly on the so-called arterial input function (AIF), that is, the concentration of contrast agent arriving in blood plasma over time. In addition, the CA is injected as bolus, due to the distance between injection area and the region of interest, and thus the AIF will not be Dirac delta function. The AIF may be measured from the DCE-MR image if a major vessel is in the field of view (Cheng, 2008). However, in many cases, the AIF cannot be measured from the image; for example, when imaging breast cancer patients. For that, in some cases, an assumed AIF from literature is applied as a bi-exponential with parameters derived following Weinmann et al. (1984) or Fritz-Hansen et al. (1996) or a mixture of the two Gaussians plus an exponential proposed by Parker et al. (2006).

The kinetic model in tissue is expressed as an ordinary differential equation, which can be solved analytically – resulting in a non-linear model for the contrast agent concentration. There are three proposed models in literature, originally proposed by Tofts and Kermode (1991), Larsson and Tofts (1992), and Brix et al. (2004). By using these models, the two kinetic parameters can be estimated from the observed contrast concentration time curve (Bender and Heinemann, 1995; Berg et al., 2014; Dikaios et al., 2014; Orton et al., 2007). In addition, a number of different techniques have been developed to quantify the estimations of two kinetic parameters. These techniques address the relationship between the estimations and the population parameters (Cheng, 2007; Gauthier, 2012). We have previously developed a method for estimating the AIF and the kinetic parameters in DCE-MRI (Farsani and Schmid, 2017). This method was developed in response to a need in the medical imaging community for the objective comparison of estimations made using different statistical methods, for example, the Bayesian Method and MLE. The main problem of our previous algorithm was the dependence of the Newton's method on the starting point, which was uniform random number.

In this paper, we have improved the previous algorithm. Our new proposal is a combination of the maximum entropy method and the optimization method using the step of λ's estimation –Teaching learning based optimization– for assessing observer performance in the classification tasks using available information when there is no concern about the random start point in the previous algorithm. This method builds on previous work on "blind" estimation of the AIF directly from tissue concentration curves (Farsani and Schmid, 2017). The maximum entropy method has proven to be an enormously powerful tool for reconstructing images from many types of data. It has a unique position as the only consistent method for combining different data into a single image. It is used most significantly in radio astronomical interferometry, where it deals routinely with images of up to a million pixels, and high dynamic range (De Martino and De Martino, 2018; Jackson et al., 2007).

The aim of this paper is to propose a novel algorithm to help more accurately estimate the parameters for the model fitting to data and, therefore, to determine more appropriate AIF with the least difference with the kinetic parameters . We performed extensive studies using empirical data to better understand the performance of our

method. In addition, a comparison was conducted among four other different estimation methods in DCE-MRI dataset, and the results of the new proposed algorithm was compared to those of the previous one (Farsani and Schmid (2017)).

A previously analyzed breast cancer data set (Farsani and Schmid, 2017; Schmid et al., 2006), provided by the Paul Strickland Scanner Centre at Mount Vernon Hospital, Northwood, UK was employed. In each scan, 46 images were acquired after administration of the contrast agent Gadolinium-DTPA every 11.9 seconds. In order to validate our method, we have designed an experiment to estimate parameters.

The rest of this paper is organized as follows: Section 2 presents the statistical formulation of the methods. Section 3 describes the data and provides the results of the mentioned methods and parameter estimation. Section 4 presents the discussion and drawn conclusions.

5.2 Methods

This study evaluates different models of arterial input functions for DCE-MRI to estimate the kinetic parameters, and compares them with previously proposed models (Farsani and Schmid, 2017). In the previous study (Farsani and Schmid, 2017), we have examined the Gamma and Exponential distributions in different situations using a maximum entropy method and a maximum a posterior Bayesian approach which was appropriate to describe the AIF. Here, we modify the previous algorithm to improve the model fit, and then, we propose several parameter estimation methods. For that, we introduce the Pharmacokinetic model for the estimation of the kinetic parameters. In the following, we explain the maximum entropy method and the proposed algorithm to find the best fitted model and, finally, the suggested model fit of AIF with its different estimation parameter techniques.

5.2.1 Pharmacokinetic Model

When the kinetic behavior of the contrast agent (CA) in the tissue (C_{tis}) of interest is considered, we use the following differential equation system:

$$\frac{dC_{tis}(t)}{dt} = K_1 C_p(t) - K_2 C_{tis}(t), \tag{5.1}$$

where the both functions $C_{tis}(t)$ and $C_p(t)$ are the concentrations of the CA at time t in the tissue of interest that is, in the extravascular-extracellular space (EES) using T1-weighted dynamic contrast-enhanced MRI (DCE-MRI), and plasma, respectively. K_1 and K_2 are the rate constants for the exchanges of CA between plasma and EES in which t depends on relaxation time values. Under $C_p(0) = 0$, Eq.(5.1) can be solved and it leads to

$$C_{tis}(t) = K_1 \int_0^t C_p(u) e^{-K_2(t-u)} du. \tag{5.2}$$

A different solution on Eq.(5.1) was presented by Murase (2004) as follows

$$C_{tis}(t) = K_1 \int_0^t C_p(u) du. - K_2 \int_0^t C_{tis}(u) du, \tag{5.3}$$

This can be written in matrix form as follows:

$$\vec{C} = \vec{A} \times \vec{K}, \tag{5.4}$$

in which $\vec{A} = \{A(1), \ldots, A(n)\}$ for $i = 1, 2, ..., n$:

$$A(i)' = (\int_0^{t_i} C_p(u)du, -\int_0^{t_i} C_{tis}(u)du), \tag{5.5}$$

$$\vec{K} = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix} \tag{5.6}$$

and

$$\vec{C} = \begin{pmatrix} C_{tis}(t_1) \\ C_{tis}(t_2) \\ \vdots \\ C_{tis}(t_n) \end{pmatrix}. \tag{5.7}$$

when two functions of time $C_{tis}(t_i)$ and $C_p(t_i)$, are measured for i=1,2,...,n, then Eq.(5.4) can be easily solved for the elements of $\vec{K}$. For that, we can rewrite Eq.(5.4) in the following form:

$$y_{tis}(t_i) = A(i)\vec{K} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \tag{5.8}$$

where $y_{tis}(t_i)$ is the observed tissue concentration at time t_i. In the previous work (Farsani and Schmid, 2017), we applied the MEM to estimate the probability density function for $\vec{C}_{tis}$ and $\vec{C}_p$ which we named them $f_{Ctis}(t)$ and $f_{Cp}(t)$, respectively. Then, we solved Eq.(5.5) and estimate the kinetic parameters K_1 and K_2 via MAP (Elfving, 1989) using Eq.(5.8). It is assumed that C_{tis} and C_p are functions of time (Farsani and Schmid, 2017) and we estimated the initial entropy probability density model for C_p for a priori model in the Bayesian framework, see Djafari and Demoment (1990).

In the present work, we have modified our previous algorithm by applying an optimization method in the parameter estimation part instead of standard Newton's method, which was effected by random starting point. The result show the estimated model is a much more better fit to data compared to the previously proposed model.

5.2.2 Maximum Entropy Method

The general form of the maximum entropy problem is to maximize the Shannon's entropy (Thomas and Cover, 2006):

$$h(X) = -\int f(x) log f(x) dx, \tag{5.9}$$

Subject to the moment constraints

$$E(\phi_k(x)) = \int \phi_k(x) f(x) dx = \mu_k, \tag{5.10}$$

where $\mu_0 = 1$, $\phi_0(x) = 1$, $\phi_k(x)$, $k = 0, ..., N$ are $N + 1$ known functions, and $\mu_k, k = 0, ..., N$ are the given expectation data. $\phi_k(x)$, $k = 0, ..., N$ can be in any functional form such as x^n,$\log(x)$, $x \log(x)$ or trigonometric and geometric functions and the main vehicles to determine the required known function is the relationship of the maximum entropy distribution with the Exponential family (Casella and Berger, 2002; Ebrahimi et al., 2008; Pougaza and Djafari, 2011). The moment constraints $\mu_1, ..., \mu_m$, are normally obtained numerically from data set using the Taylor's theorem (Casella and Berger, 2002). Using an appropriate optimization method, where the objective function is Shannon's entropy, the general form of the maximized density is obtained as follows (Thomas and Cover, 2006):

$$f(x) = e^{-\sum_{k=0}^{N} \lambda_k \phi_k(x)}, \quad x \in S, \tag{5.11}$$

where λ_k should be chosen such that $f(x)$ in Eq.(6.25) satisfies the known moment constraints in Eq.(6.24). The parameters $\lambda = [\lambda_0, ..., \lambda_N]$ are calculated to find the class of the maximum entropy distributions. To obtain the $N+1$ unknown parameters, the following set of $N + 1$ nonlinear equations is solved ($1 \leq k \leq m$):

$$G_k(\lambda) = \int \phi_k(x) e^{-\sum_k \lambda_k \phi_k(x)} dx = \mu_k. \tag{5.12}$$

To solve Eq.(5.12) using the proposed *modified algorithm*, we suggest the Teaching learning based optimization (TLBO) which resolves the problem of random starting points. Instead, it can measure the mean of all possible parameter estimations to fit a better model to data.

Maximum A Posterior Probability Approach

The Bayesian approach Maximum a Posteriori (MAP) prepared a common basis in statistical methods to solve ill-posed image reconstruction problems. The main thing is that, solutions of MAP are highly dependent on a priori model. The proposed methods developed in literature are based on prior models that describe the properties of the expected image rather than the properties of the studied object. The main difficulty of the Bayesian approach is the determination of the a priori probability law for the image and determining its parameters from the data. In this situation, there are two different classes of imaging problems: restoration of distorted images and tomographic reconstruction of objects. When solving the image restoration problems, it is usually impossible to include a priori information about real objects. A priori information refers to the expected image to be restored. In contrast, when reconstructing tomographic images, one can often propose the physical model of the object to be reconstructed and determine a priori information by using this model. Currently, in tomographic problems, one uses a priori information, which was initially developed to solve the image restoration problems (Denisova, 2019; Sparavigna, 2019).

In this paper, we discuss these approaches to specifying a priori probability in both the image restoration problem and the image reconstruction problem. We consider the most widely used forms of a priori information which named entropy-based prior, based on the entropy principle, was successfully applied in the fields of X-ray-, radio- and

gamma-astronomy, plasma tomography (Denisova, 2019; Mohammad-Djafari, 1996; Skilling, 1988; Sparavigna, 2019).

To estimate the kinetic parameters, considering the general form of Eq.(5.8) as follows (Djafari and Demoment, 1990):

$$y = Ax + b, \tag{5.13}$$

where $\vec{x}$ is an unknown positive vector, $\vec{y}$ is a vector of measurement with a linear transformation A and b presents the noise measurement, which is supposed to be uncorrelated normal distribution with zero-mean and additive. To determine the unknown vector $\vec{x}$, we apply the Bayes rule. We would like to find the peak of the conditional distribution for x given y, $f(y|x)$, and take the corresponding x-vector as a solution of reconstruction problem. This is a well-known method in parameter estimation named maximum a posteriori (MAP). This method requires the specification of $f(x)$ and $f(y|x)$. The estimator $\hat{x}$ maximizes the posterior distribution obtained by Bayesian formula (Elfving, 1989):

$$f(x|y) = f(y|x)f(x)/f(y), \tag{5.14}$$

In Eq.(5.14), $f(y)$ is independent of x, $f(y|x)$ is in fact related to the noise probability law and $f(x)$ is a prior law on x. With a linear relation between x and y, and assuming uncorrelated normally distributed errors in the linear relation by knowing the only variance σ^2 of the noise, it is easy to obtain $f(y|x)$ as a correlated normally distribution via MEM as follows:

$$\begin{aligned} f(y|x) &\approx \exp[-T(x)], \\ T(x) &= [y - Ax]^t[y - Ax]/\sigma^2. \end{aligned}$$

A possibility to choose a priori distribution $f(x)$ is again the MEM where the general model belongs to the Exponential family. The advantage in using this concept is to be most objective or maximally uncommitted with respect to missing information Elfving (1989). Then the solution of our problem is the mode of the posterior probability distribution as follows:

$$\begin{aligned} \hat{x} &= Arg\max_{x>0} f(x|y) \\ &= Arg\max_{x>0}\{f(y|x)f(x)\} \end{aligned} \tag{5.15}$$

Weibul Distribution

In the literature, several probability distributions have been proposed as AIF, such as Gamma, mixture of Gaussian and Exponential distribution, and bi-Exponential distributions (Fritz-Hansen et al., 1996; Parker et al., 2006; Weinmann et al., 1984). Using the proposed algorithm, we propose the Weibull distribution to investigate how well it can fit to the data.

The Weibull distribution (named after the Swedish physicist Weibull, who applied it when studying material in tension and fatigue in the 1930s) provides a close approximation to the probability laws of many natural phenomena. The Weibull probability

density function, a two parameter function, is expressed mathematically as

$$f(x) = \frac{k}{c^k} x^{k-1} e^{-(\frac{x}{c})^k}, \tag{5.16}$$

and the cumulative distribution function is

$$F(x) = 1 - e^{-(\frac{x}{c})^k}, \tag{5.17}$$

where k is the shape parameter and c is the scale parameter in Bain and Antle (1967); Stevens and Smulders (1979).

Determination of Parameters

The parameters of Weibull distribution (k and c) can be found by a number of ways. Some of them are mentioned here.

Methods of Moments

The method of moments is considered as an alternative to maximum likelihood method. The first two moments of the Weibull density function are utilized to calculate the parameters k and c. The calculations are based on standard deviation, average and gamma function for parameter $(1 + 1/k)$. This method is suggested by Justus et al. (1978). The two moments of the distribution are given in Eq.(5.18) which help in calculating shape and scale parameters. The sample mean and standard error are

$$\begin{cases} \bar{x} = c\Gamma(1 + 1/k) \\ \sigma = c(\Gamma(1 + 1/k) - \Gamma^2(1 + 1/k))^{1/2}, \end{cases} \tag{5.18}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function.

Empirical Measurement Method

The Empirical method is the special case of the moment method (Akdağ and Dinler, 2009; Justus et al., 1978; Morgan et al., 2011):

$$\begin{cases} k = (\frac{\sigma}{\bar{x}})^{-1.086}, \\ c = \frac{\bar{x}}{\Gamma(1+\frac{1}{k})}, \end{cases} \tag{5.19}$$

where σ is the sample standard deviation.

Maximum Likelihood Method

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model; given observations. The method obtains the parameter estimates by finding the parameter values that maximize the likelihood function. In frequentest inference, MLE is one of several methods to get estimates of parameters without using prior distributions. The Weibull distribution can be fitted to time-series

data using the maximum likelihood method as suggested by Stevens and Smulders (1979).

In here, x_i is the data in step i and n is the number of non-zero data points, the shape parameter k and the scale parameter c are estimated using the following two equations:

$$\begin{cases} k = \left(\frac{\sum_{i=1}^{n} x_i^k \ln(x_i)}{\sum_{i=1}^{n} x_i^k} - \frac{\sum_{i=1}^{n} \ln(x_i)}{n} \right)^{-1}, \\ c = (\frac{1}{n} \sum_{i=1}^{n} x_i^k)^{1/k}, \end{cases} \tag{5.20}$$

k can be solved iteratively in the first part of Eq.(5.20), after which the second part of Eq.(5.20) can be solved analytically to get c.

Modified Maximum Likelihood Method

When the data is available in the form of the frequency distribution, we apply the modified maximum likelihood method.

$$\begin{cases} k = (\frac{\sum_{i=1}^{n} {x_i}^k \ln(x_i) P(x_i)}{\sum_{i=1}^{n} {x_i}^k P(x_i)} - \frac{\sum_{i=1}^{n} \ln(x_i)}{P(x \geq 0)})^{-1}, \\ c = (\frac{1}{P(x \geq 0)} \sum_{i=1}^{n} {x_i}^k P(x_i))^{1/k}, \end{cases} \tag{5.21}$$

where n represents the Weibull frequency, $P(x \geq 0)$ is the probability that the random variable equals or exceeds zero. In Eq.(5.21), k must be solved iteratively, after which c can be solved explicitly (Yang et al., 2004).

Non-linear Least Squares (Non-LS) Method

The Non-LS method has many similarities to the linear LS method. The observed data are also sorted in an ascending fashion, and subsequently paired with the failure probabilities, obtained by the estimators. It differs from the LLS method as a non-linear regression, using a Gauss–Newton algorithm, is directly carried out to achieve the best fitted curve of a Weibull function. This method was used to estimate Weibull parameters in some other fields, but has not been applied in the Weibull estimation of castings and brittle materials (Li et al., 2017).

5.2.3 Teaching-learning-based optimization (TLBO)

In the optimization of a design, the design objective could simply be to minimize the cost of production or to maximize the efficiency of production. An optimization algorithm is a procedure which is executed iteratively by comparing various solutions until an optimum or a satisfactory solution is found. Teaching-learning-based optimization (TLBO) is a population-based metaheuristic search algorithm inspired by the teaching and learning process in a classroom proposed by Rao et al. (2011). It has been successfully applied to many scientific and engineering applications in the past few years. In the basic TLBO and most of its variants, all the learners have the same probability of getting knowledge from others. However, in the real world, learners are different, and the learners' learning enthusiasm are not the same, resulting in different probabilities of acquiring knowledge. TLBO utilizes two productive operators, namely, teaching phase

and learning phase to search for good solutions. Due to its attractive characteristics such as a simple concept, lack of specific algorithm parameters, easy implementation, and rapid convergence, TLBO has captured great attention and been extended to handle constrained, multi objective, large-scale, and dynamic optimization problems. In the present work, we apply an optimization method to modify the previous algorithm to estimate the maximum entropy model. Furthermore, to determine the parameters of the estimated model, we propose different parameter estimation methods (Rao et al., 2011, 2012).

Implementation

We have implemented a MATLAB code for the proposed algorithm and all calculations were done using MATLAB. To make our computations easier, we suggest to estimate a probability distribution model via the MEM for $\vec{C}_{tis}$ and $\vec{C}_p$. In this procedure, we have the following steps:

(1) Determining $\phi_k(x)$ and computing their expectations numerically from the data set based on Taylor's theorem (Thomas and Cover, 2006).

(2) Using the appropriate ***optimization method*** in order to determine the unknown function with the Shannon entropy as target function. The general form is in Eq.(6.25), ($f_{C_{tis}}(t)$ and $f_{C_p}(t)$),

(3) Applying the proposed method to find λ_k such that $f(x)$ (Eq.(6.25)) satisfies the known moment constraints (Eq.(6.24)), ($\hat{f}_{C_{tis}(t)}$ and $\hat{f}_{C_p}(t)$),

(4) Estimating the kinetic parameters $\vec{K}$, we replace $\hat{f}_{C_{tis}}(t)$ and $\hat{f}_{C_p}(t)$ in Eq.(5.8) and solve it,

(5) Applying the Kullback-Leibler divergence $D_{K-L}(f||g)$ to check the accuracy of the final AIF, $\hat{f}_{C_p}(t)$ in comparison to the empirical distribution of data $g(C_p)$,

$$D_{K-L}(\hat{f}||g) = \int_s \hat{f}_{C_p}(t) log \frac{\hat{f}_{C_p}(t)}{g(C_p)} dt. \tag{5.22}$$

(6) With the predicted values $\hat{x}_1, ..., \hat{x}_m$ and the observed values $x_1, ..., x_m$:

$$RMSE = \Big[\frac{1}{m}\sum_{i=1}^{m}(x_i - \hat{x}_i)^2\Big]^{1/2}, \tag{5.23}$$

$$\chi^2 = \frac{\sum_{i=1}^{m}(x_i - \hat{x}_i)^2}{m-n}, \tag{5.24}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \tag{5.25}$$

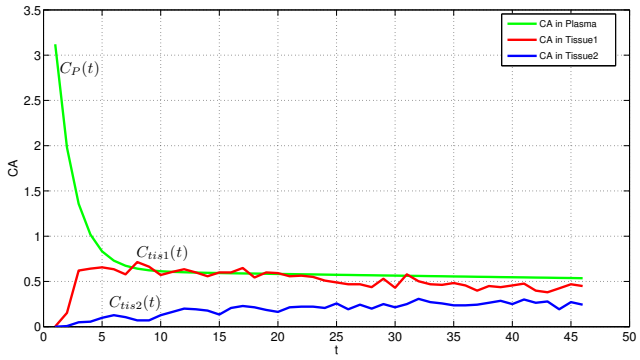

Figure 5.1: Plot of Contrast Agent (CA) in Plasma ($C_p(t)$) and Tissue ($C_{tis}(t)$) for both Patients 1 & 2

5.3 Data Application

5.3.1 Data description

We use a previously analyzed breast cancer data set (Schmid et al., 2006) provided by Paul Strickland Scanner Centre at Mount Vernon Hospital, Northwood, UK. Here, we analyze pre-treatment DCE-MRI scans of two patients. For each scan, 46 images were acquired after administration of the contrast agent Gadolinium-DTPA every 11.9 seconds. The imaging parameters of the T1-weighted FLASH images were TR=11 ms, TE=4.7ms, α=35, the parameters of the proton density-weighted image were TR=350 ms, TE=4.7 ms, α=6. Field of view was the same for all scans, $260 \times 260 \times 8$ mm per slice, so voxel dimensions are 1.016×1.0168 mm. A scan consists of three sequential slices of 256×256 voxels and one slice placed in the contralateral breast as control, which we do not use for our analysis. A dose of D = 0.1 mmol per kg body weight of Gd-DTPA was injected after the fourth scan using a power injector with 4 mL/sec with a 20 mL saline flush also at 4 mL/sec.

In order to calculate contrast agent concentration $C_t(t)$, the signal intensity is converted to T_1 relaxation time values using T_1-weighted images, proton density weighted images and data from calibration phantoms with known T_1 relaxation times. The Gd-DTPA concentration can then be computed via $C_t(t) = \frac{1}{r_1}\left[\frac{1}{T_1(t)} - \frac{1}{T_{10}}\right]$, where T_{10} is the T_1 value without contrast, computed as mean value of the first four images, and $r_1 = 4.24$l/s/mmol is the longitudinal relativity of protons *in vivo* due to Gd-DTPA. As initial assumption for $\vec{C}_p = [C_p(t_1), ..., C_p(t_{46})]$ is taken from Fritz-Hansen et al. (1996). The scans were performed before treatment. After a total of 18 weeks of chemotherapy (5-fluorouracil, epirubicin, and cyclophosphamide) pathological response to treatment was assessed (Farsani and Schmid, 2017). Fig.5.1 shows the frequency model of data for two patients $\vec{C}_{tis-1}$, $\vec{C}_{tis-2}$ and $\vec{C}_p$ subject to the time t. This is the primary model for data $C_p(t)$ changing the inverse problem to forward.

5.3.2 Results for One Patient

In the previous study, we have evaluated the performance of the MEM combined with the Newton's method in a Bayesian framework to estimate the best fitted model to CA data. Here, we have modified the previous algorithm. In the modified algorithm, we have examined the optimization method with a different class of constraints to get better results. Among them, the following constraints which can fit the Weibull distribution, made the results more appropriate than the previous method (Fig.5.2, Fig.5.3). For $\vec{C_p}(t)$

$$\begin{cases} \int_t f_{C_p}(t)dt = 1, \\ \int_t \log(C_p) f_{C_p}(t)dt \ = -0.4465, \\ \int_t C_p^3 f_{C_p}(t)dt \ = 1.0930 \end{cases} \tag{5.26}$$

the general form of the resulting priori probability model $\hat{f}_{C_p}(t)$ is the Weibull distribution as follows:

$$f_{C_p}(t) = e^{-\lambda_0 - \lambda_1 \log(C_p(t)) - \lambda_2 C_p^3(t)}, \tag{5.27}$$

where the final ME multipliers λ's and the Weibull parameters are estimated as follows

$$\hat{f}_{C_p}(t) = \exp(-0.7466 - 1.4944 \log(t) - 0.1128 t^3) + 0.5. \tag{5.28}$$

and based on the ME form of Eq.(5.16)

$$f(x) = e^{\log(\frac{k}{c^k}) + (k-1)\log(x) - (\frac{x}{c})^k}. \tag{5.29}$$

in which

$$\begin{cases} \lambda_0 = -\log(\frac{k}{c^k}), \\ \lambda_1 = -(k-1), \\ \lambda_2 = c^{-k}. \end{cases} \tag{5.30}$$

Then, according to Eq.(5.28) and Eq.(5.30), the Weibull parameters will be $c = 1.8498$, $k = 3$ where the mean of absolute error, D_{K-L} divergence and entropy are 0.0470, 0.0438 and 0.2021 respectively, (Fig.5.2).

Table 5.1 and Fig.5.4 show the results of different models (Section.5.2.2) using the new algorithm. In each case, we have applied evaluation methods to check the accuracy of the estimated model. Furthermore, the high measure of entropy shows the more superiority of the selected probability distribution to fit data. Based on the Jaynes's principle (Jaynes (1957a)), the concept of MEM is to find the distribution that maximizes the entropy. In addition, $\hat{f}_{C_p}(t)$ is the modified ME probability distribution for AIF. In comparison to the previous study, We have estimated a model which has the maximum entropy (Fig.5.3, Fig.5.4).

Table 5.2 includes the Weibull parameters estimated via different estimation methods mentioned in (Section.5.2.2). All estimation methods are presented in Fig.5.5. Then, we compared the accuracy of the results using: Empirical Method (EM), Maximum Likelihood (ML), and the Modified Maximum Likelihood Method (MMLM). Table 5.2 includes the parameter estimation via different methods for $\vec{C_p}$. The main results are presented in Table 5.3. There are four evaluation measures to investigate

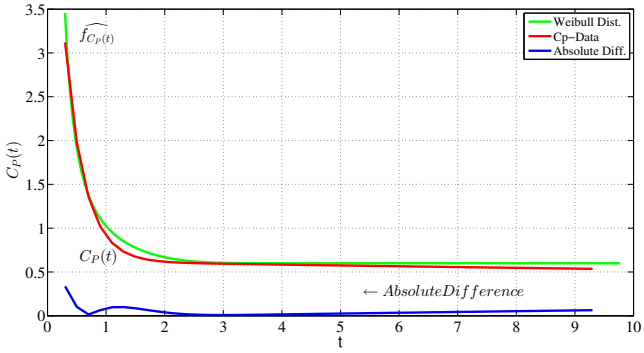

Figure 5.2: Modified Maximum Entropy Estimated Input Function, $k = 3$ & Literature Input Function

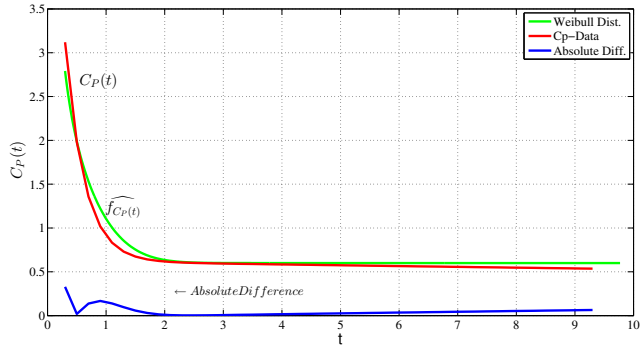

Figure 5.3: Modified Maximum Entropy Estimated Input Function, $k = 2$ & Literature Input Function

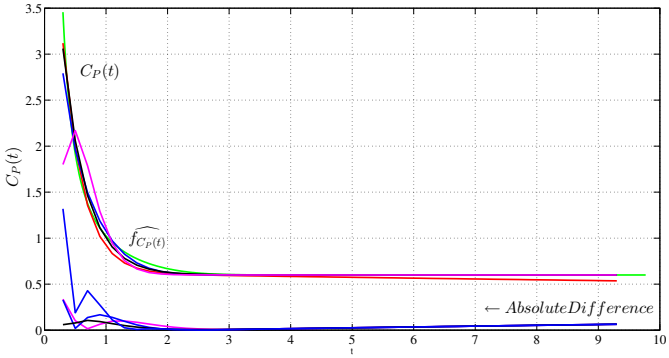

Figure 5.4: Modified Maximum Entropy Estimated Input Function & Literature Input Function

Estimated Distribution	MAE	D_{K-L}	Entropy
Gamma	0.0775	0.0285	0.0303
Exponential	0.0375	0.0363	0.0872
Weibull ($k = 3$)	0.0470	0.0438	0.2026
Weibull ($k = 2.6$)	0.0403	0.0389	0.1755
Weibull ($k = 2.5$)	0.0409	0.0394	0.1812
Weibull ($k = 2.4$)	0.0414	0.0399	0.1871
Weibull ($k = 2.3$)	0.0420	0.0404	0.1929
Weibull ($k = 2.2$)	0.0427	0.0410	0.1989
Weibull ($k = 2$)	0.0471	0.0342	0.1471

Table 5.1: Comparison of the Modified Maximum Entropy Estimated Input Function and the MEM

Methods	K	C
Empirical Measurement (EM)	1.6469	0.7787
Method of Moments (MOM)	1.9125	0.7850
Maximum Likelihood Method (MLE)	1.8005	0.7890
Modified Maximum Likelihood Method (MMLE)	2.0201	0.7758
Least squares method (LSM)	2.7767	0.7518
Modified Maximum Entropy Approach (MMEM)	2	1.3765
Modified Maximum Entropy Approach (MMEM)	2.2	1.8198
Modified Maximum Entropy Approach (MMEM)	2.3	1.7983
Modified Maximum Entropy Approach (MMEM)	2.4	1.7774
Modified Maximum Entropy Approach (MMEM)	2.5	1.7573
Modified Maximum Entropy Approach (MMEM)	2.6	1.7380
Modified Maximum Entropy Approach (MMEM)	3	1.8498

Table 5.2: Weibull Parameters Estimation Methods

Methods	***RMSE***	***Chi-Square***	R^2	***Adjust*** R^2
EM	0.286	0.0755	0.631	0.622
MOM	0.255	0.0691	0.67	0.663
MLE	0.278	0.1191	0.57	0.58
MMLE	0.274	0.0771	0.636	0.628
LSM	0.194	0.2854	0.535	0.525
MMEM, $k = 2$	0.051	0.0109	0.984	0.984
MMEM, $k = 2.2$	0.0338	0.0029	0.995	0.994
MMEM, $k = 2.3$	0.0323	0.0020	0.995	0.995
MMEM, $k = 2.4$	0.0315	0.0013	0.995	0.995
MMEM, $k = 2.5$	0.0315	9.2214e-04	0.995	0.995
MMEM, $k = 2.6$	0.0320	7.5687e-04	0.995	0.995
MMEM, $k = 3$	0.048	0.0086	0.990	0.990

Table 5.3: Evaluating Methods & the Results

how the algorithm perform. For each method, four measurements are available, among which, the modified maximum entropy method has the best fit to the data.

Based on the results of Table 5.1 to Table 5.3, in the cases of the Weibull model with $k = 2.2$ and $k = 2.3$, we have achieved a much better fit model to the data. Actually, in Table 5.2 we have examined the the parameter estimation of the Weibull distribution via different methods in comparison to those via MMEM with different k to see in which case the estimated model fits better to the data. Then, in the next table (Table 5.3), we evaluated the results based on some evaluation measurements such as root mean square error ($RMSE$), goodness of fit (χ^2), determination coefficient (R^2) and adjust determination coefficient (R^2) which show that using the MMEM leads to a better fit. The proposed algorithm gives the estimation with the lowest absolute error and D_{K-L} divergence. The most important point of the results is that the proposed algorithm gives the model with highest measure of entropy.

5.3.3 Results for the Whole Study

To further evaluate the proposed algorithm and the fit of the estimation model to the data, we analyzed the data of 12 patients in total. MMEM/MAP was utilized to estimate the kinetic parameters, and the results were compared to the results of the previously proposed methods. Fig. 5.6 shows data and the literature AIF, respectively. In addition, Fig. 5.8 shows the Logarithm of absolute error for each patient. The difference between assumed AIF and estimated AIF using the modified Maximum Entropy Method is obvious. In particular, the input function in the first two minutes is typically estimated higher than the assumed AIF, whereas there is negligible difference after about two and a half minutes. However, the correct estimation of the AIF at the onset is the most important for the correct estimation of the kinetic parameters. Kullback-Leibler divergence values change from 0.001 to 0.0637 for all the twelve patients.

For all the patients the results are presented in Table 5.4. The Weibull models for

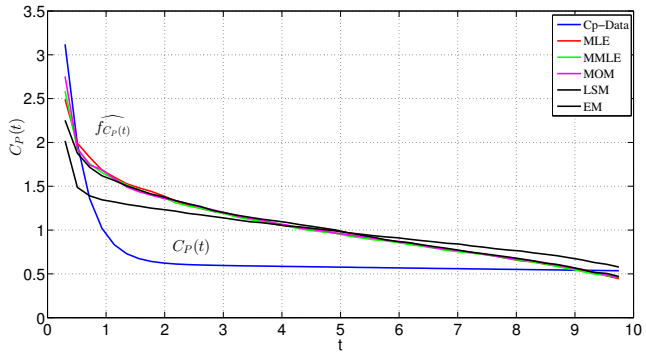

Figure 5.5: Fitting the Model on C_p-Data via Different Parametric Methods

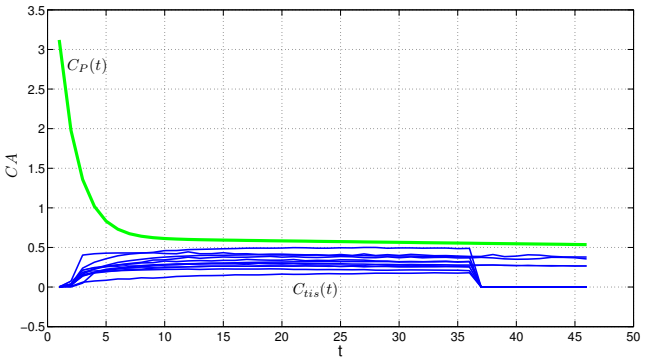

Figure 5.6: Plot of Contrast Agent (CA) in Plasma ($C_p(t)$) and Tissue ($C_{tis}(t)$) for 12 Patients

$k = 2.2$ to $k = 2.4$ have the best fit to the data based on all evaluated measurements. In addition, Fig. 5.7 shows the estimated kinetic parameter k_1 using MEM/MAP, MEMM/MAP and assumed AIF/ML methods.

5.4 Discussion & Conclusions

Here, we proposed an algorithm to determine the final AIF and then estimate the kinetic parameters in the DCE-MRI data in cases with no input function. For that, we modified our previous algorithm by the optimization method. Several methods were proposed for estimating the parameters of the Weibull distribution in the data analysis of the kinetic models. Then the method of moments, the maximum likelihood method, the empirical method, and others have been employed to calculate the parameters 'k' and 'c' of the estimated distribution function. The data from the previous study were used to investigate the accuracy of these models for determining of Weibull parameters. The application of each method was demonstrated using a sample data set, and to determine the accuracy of each method, the measured data obtained from literature were compared with our data (Table 5.2 and Table 5.4). Various tests were used to

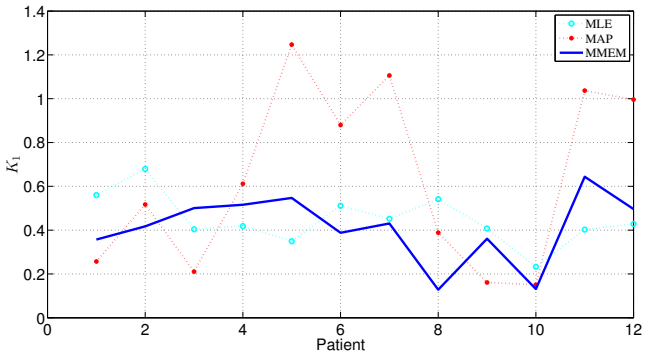

Figure 5.7: $k1$ estimated using MMEM/MAP and assumed AIF/ML & MEM/MAP for all 12 patients

Patient	1	2	3	4	5	6
k_1	0.1637	0.1016	0.7175	0.1650	0.5959	1.0477
k_2	0.0210	0.3688	0.1073	0.2079	0.1233	0.0072
Patient	7	8	9	10	11	12
k_1	0.6309	0.7980	0.1085	0.4327	0.544	1.0225
k_2	0.0701	0.3861	0.2377	0.0839	0.235	0.0271

Table 5.4: Kinetic Parameters Estimation via the Modified Maximum Entropy Method for 12 patients

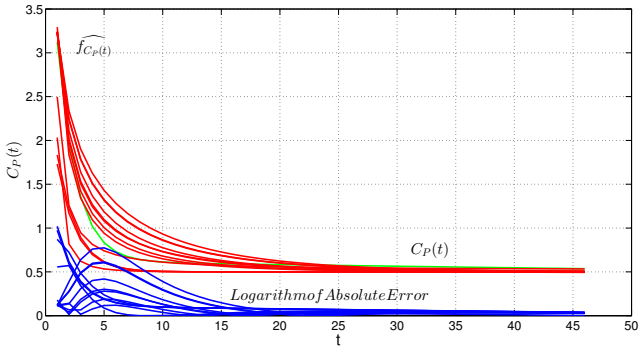

Figure 5.8: Modified Maximum entropy Estimated AIF for 12 patients, literature input function

analyze the accuracy of the compared methods (Table 5.3).

According to the results, it can be concluded that the suitability of these methods may vary with the sample data such as data size, sample data distribution, sample data format and of fit tests. The MMEM is an efficient method to fit Weibull distribution and determine the k and c parameters. This fact is also supported by means of the Fig.5.8, and Table 5.2. It is also observed from the statistical analysis that the values of the RMSE, χ^2 tests have magnitudes close to each other for the MMEM.

Chapter 6

Statictical Method for Quantitative Measurement of Colocalization

Abstract

Accurately localizing molecules within the cell is one of main tasks of modern biology, and colocalization analysis is one of its principal and most often used tools. Despite this popularity, interpretation is often uncertain because colocalization between two or more images is rarely analyzed to determine whether the observed values could have occurred by chance. We link the concept of dependency in the structure of colocalization in two or more channels to an important problem in statistics which is to estimate a joint probability distribution from its marginals. We introduce a entropy-based statistical method that quantifies, the amount of colocalization of two fluorescent-labeled proteins in an image via different class of copula. The new optimization method have been proposed to estimate the maximum entropy marginals of both red and green channels and it can be combined with copula to measure colocalization even when pixels do not show any statistical correlation. Using our methodology, we are able to determine not only whether the labeled molecules colocalize with a probability greater than chance, but also it can measure amount of anti-colocalization even in high background settings. The proposed models are categorized into three groups based on the class of copula, the univariate maximum entropy model of both channels and their model structures. Then, a numerical example is presented to illustrate the formulation and implementation of each type of the entropy copula model. In addition, the potential application of the maximum entropy copula in the fluorescence microscopy images data are discussed. Finally, the results of different classes of copula are compared to get the best fitted model. More generally, this algorithm can be used to answer a variety of biological questions involving protein-protein interactions or co-compartmentalization and can be generalized to colocalization of more than two colors.

keywords: Medical Imaging; Gaussian Copula; Fluorescence microscopy Image; Kendal's τ; Maximum Entropy Method; Colocalization

6.1 Introduction

Data analysis is crucial in almost every field of research such as genomics, economics, physics, medical, social, and political sciences. Identifying associations between/among variables is often required in analysis of large datasets (Hastie et al., 2009). It is common to have many variables in a dataset and it is difficult to manually examine the relation between each pair of variables (Reshef et al., 2011). It is also difficult to identify the important variables if the correlation among them is not discovered. There are several different measures to quantify the association between variables in a dataset including Pearson's correlation, Spearman's correlation, distance correlation, maximal information coefficient (MIC), maximal correlation, and mutual information. Some of these correlation measures can only detect linearly correlated data such as the well-known Pearson's correlation while some measures can also detect nonlinear correlation such as maximal correlation and MIC. In addition, some correlation measures can characterize the independence. This means if the correlation score yields a value of zero, one can conclude that the two variables are independent Adler and Parmryd (2010); Aggarwal and Gupta (2019); Almaraz-Damian et al. (2020); Benesty et al. (2009); Deebani and Kachouie (2018); Mukaka (2012); Peng et al. (2020); Schober et al. (2018); Thao et al. (2019); Zhao et al. (2019).

In more general view, considering the concept of connection between the multivariate and univariate function, there is an increasing interest concerning copula. Copula appears to be a powerful tool to model the structure of dependence (Huang and Emura, 2019; Ouyang et al., 2009; Zhang et al., 2008). The word copula originates from the Latin meaning link, chain, union. In statistical literature, according to the seminal result in the copula's theory stated by Sklar (1959), a copula is a function that connects a multivariate distribution function to its univariate marginal distributions. There is an increasing interest concerning copulas, widely used in Financial Mathematics and in modeling of Environmental Data (Genest and Favre, 2007; Joe, 1994). Recently, in Computational Biology, copulas were used for DNA analysis (Kim et al., 2008). Copulas are useful for constructing joint distributions, particularly with non-Gaussian random variables (Joe, 1997).

In addition, the text by Nelsen (1999) is a comprehensive reference for the general theory of copulas. Much of the work on copulas has been related to construction of general families with desirable properties. Although copulas have been used in one form or another for many years much of the theory has been developed relatively recently (Borwein et al., 1994; Chen et al., 2006; Genest et al., 1995, 2009).

By considering an important problem in statistics which is to determine a joint probability distribution from its marginals and the concept of copula which is a function that connects a multivariate distribution function to its univariate marginal distributions, a number of methods to construct a joint distribution have been proposed and applied. These methods include the kernel density estimation, entropy method, and copulas, to name a few (Dunn et al., 2011). The entropy–copula methods generally retain the advantage of the commonly used parametric copula: the joint distribution construction will be independent of the marginal distributions. These methods look to be powerful techniques to apply them for, modeling biological variables (Aghakouchak,

2014).

In the biological context, the accurate intracellular localization of proteins to their specific compartments is a rich source of information for the study of biological processes and it is one of main tasks of modern biology, and colocalization analysis is one of its principal and most often used tools (Kalaidzidis et al., 2015). Colocalization happened when two different molecules attach to the same – or nearby –structures within the cell to fulfill a biological function. In contrast, anti-colocalization describes a system where different proteins are not present at the same location at the same time (Betzig et al., 2006; Garini et al., 2006; Helmuth et al., 2010; Hess et al., 2006; Kalaidzidis et al., 2015; Kobayashi et al., 2009; Manders et al., 1993; Wang et al., 2017).

Despite this popularity, interpretation is often uncertain because colocalization between two or more images is rarely analyzed to determine whether the observed values could have occurred by chance. In addition, colocalization between two fluorescent labeled molecular species (typically between proteins) is a common question in optical microscopy. However, existing colocalization techniques are generally visual-based and therefore highly prone to random error and bias (). We link the concept of dependency in the structure of colocalization problem from two or more channels to an important problem in statistics which is to estimate a joint probability distribution from its marginals ().

Fig. 6.1 shows a two-channel fluorescence microscopy image with markers for γ-H2AX (red) and H3K9me2 (green) Seiler et al. (2011). The zoom up (Fig. 6.1 middle) depicts a part of the cell, where both red and green signals overlap, producing yellow spots. However, in other areas red and green signal are separated. Overall, γ-H2AX and H3K9me2 are know to be independent, hence we expect no colocalization of the markers.

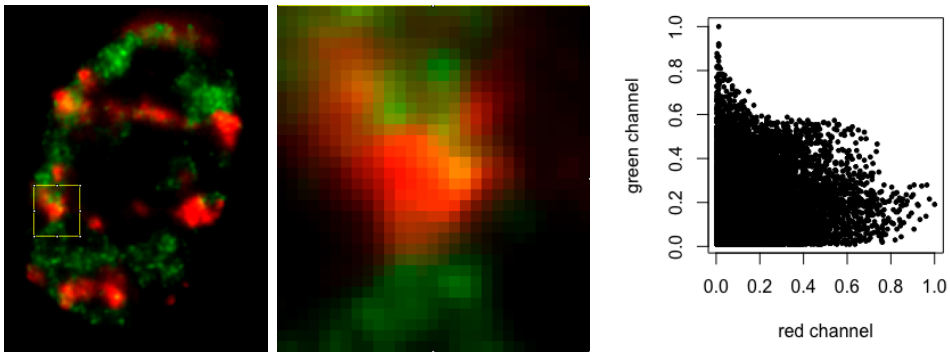

Figure 6.1: Two-channel image of γ-H2AX/H3K9me2 image. Left: Central slice. Middle: zoom on detail (yellow box in left figure). Left: scatter plot of red and green channels from all slices.

Tab. 6.1 shows the correlation coefficients between the red and green signal with different levels of threshold. For low thresholds, a significant positive correlation is found. With larger thresholds, that is, with suppression of background, the positive correlation disappears. However, if the threshold is to large, we get a negative correlation which is significant at a 1% level. So we see, that the threshold has a very big

threshold	no	0	0.05	0.1
correlation	0.338	0.0856	0.0761	0.0405
p-value	$< 2.2 \cdot 10^{-16}$	$< 2.2 \cdot 10^{-16}$	$< 2.2 \cdot 10^{-16}$	$9.1 \cdot 10^{-5}$
threshold	0.2	0.3	0.4	0.5
correlation	−0.0822	−0.0137	−0.139	−0.412
p-value	$6.3 \cdot 10^{-5}$	0.755	0.116	0.0089

Table 6.1: Correlation Coefficient and Significance Test Results of $\rho = 0$ & significance level $\alpha = 0.05$ between two channels of H3K9me2.

impact on the results of a correlation analysis.

Of particular interest in fluorescence microscopy imaging is the so-called colocalization analysis (Helmuth et al., 2010; Kobayashi et al., 2009; Ronneberger et al., 2008). Several approaches have been proposed for measuring colocalization using cross-correlation analysis (Barbaresco and Chevalier, 2008; Hussain, 2009; Manders et al., 1993) or cluster analysis of the two-dimensional histogram (Demandolx and Davoust, 1997; Garini et al., 2006; Valeur and Berberan-Santos, 2012). However, most of those approaches are either qualitative (only proving existence of colocalization, e.g., cross correlation analysis) or subjective (due to manual identification of clusters in two-dimensional histograms). The most commonly used quantitative measures for colocalization are Pearson's correlation coefficient and Manders' split coefficients (Manders et al., 1992). Pearson's correlation coefficient was first introduced to the microscopy community (Fletcher et al., 2010). It measures the linear relationship of the intensities between the two channels, and a strong correlation indicates that a large intensity in one channel is often associated with a large intensity in the other. Another popular colocalization measure is the Manders' split coefficients proposed by (Dunn et al., 2011; Fletcher et al., 2010; Wang et al., 2017). These coefficients measure the fractions of signal in one channel that overlap with the other. For both methods there are some complexities within biological contexts when measuring colocalization. For that, there is an attempt to consider a more robust method to quantify more general positive dependencies between two probes. For instance, the idea of nonparametric correlation coefficient in colocalization analysis.

In this study, the aim is to measure colocalization with considering an ill-posed inverse problem which is to estimate the joint probability distribution from it's channels. In the bivariate case, the marginal probability density functions $f_1(x)$ and $f_2(y)$ are related to their joint distribution $f(x, y)$ via horizontal and vertical line integrals. Interestingly, this is also the case of a very limited angle $X - ray$ CT problem where $f(x, y)$ is an image representing the distribution of the material density and $f_1(x)$, $f_2(y)$ are the horizontal and vertical line integrals (Pougaza et al., 2010). The marginals are the univariate maximum entropy distribution of the red and green signals. In addition, we modified the previous algorithm by adding a new optimization method in the parameter estimation step. Despite the previous study (Farsani and Schmid, 2020) in which we have considered the Gaussian copula to determine the dependence structure of the red and green channel based on simulated and empirical data, in here the proposed copula models are categorized into different groups based on their model structures. Then,

a simple numerical example is used to illustrate the formulation and implementation of each type of the model. Finally, the empirical application in cell biology and the fluorescence microscopy is provided to evaluate the proposed methods which followed by an example application of the entropy–copula concept to colocalization analysis.

The remainder of this paper is outlined as follows. The entropy and copula and maximum entropy copula are introduced in sections 2, respectively. Data application via the recently developed entropy– copula methods are reviewed and categorized, followed by a numerical example in section 4. The last section, summarizes the results and conclusions.

6.2 Methods

6.2.1 Joint Probability Distribution with Copula

Copulas are used to describe the dependence between random variables. Fisher Fisher (1997) gave two major reasons as to why copulas are of interest to statisticians: "Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions." According to Sklar's theorem (Rüschendorf, 2013) there exists a copula which relates the marginal distributions yielding to the joint distribution. The problem then becomes the choice of a copula. Note that there are many other ways to derive families of continuous multivariate distributions with given univariate marginals (Genest and MacKay, 1986; Genest and Rivest, 1993; Pougaza and Djafari, 2011; Rüschendorf, 2013). By $F(x, y)$ we denote a continuous bivariate cumulative distribution function (CDF), and $f(x, y)$ its bivariate probability density function (PDF). Let $F_1(x)$, $F_2(y)$ be the marginal CDF's and $f_1(x)$, $f_2(y)$ their respective PDF's. A bivariate copula C is a function from I^2 to I with the following properties:

- $\forall u, v \in I, C(u, 0) = 0 = C(0, v)$,
- $\forall u, v \in I, C(u, 1) = u \; and \; C(1, v) = v$,
- $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all $u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2, v_1 \leq v_2$.

One can construct copulas C from joint distribution functions by

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \tag{6.1}$$

where the quantile function is $F_i^{-1}(t) = inf\{u : F_i(u) \geq t\}$, (Pougaza and Djafari, 2011; Romeo et al., 2006). For any multivariate absolutely continuous distribution, with CDF F and marginal CDF's F_i, the copula C is such distribution function on $(0; 1)^p$ with uniform one-dimensional marginals that it holds

$$F(x_1, ..., x_p) = C(F_1(x_1), ..., F_p(x_p)). \tag{6.2}$$

With the copula density c defined by

$$c = \frac{\partial^p C}{\partial F_1 ... \partial F_p} \tag{6.3}$$

the joint density can be expressed as

$$f(x) = c(F_1(x_1), ..., F_p(x_p)) \prod_{i=1}^{p} f_i(x_i). \tag{6.4}$$

Sklar's Theorem

Let F be a joint distribution function with marginals F_1 and F_2. Then, there exists a copula C such that, for all $x, y \in (-\infty, \infty)$

$$F(x, y) = C(F_1(x), F_2(y))$$

. If F_1 and F_2 are continuous, then the copula C is unique; otherwise, C is uniquely determined on (Range of F_1 and F_2). Conversely, if C is a copula and F and G are univariate distribution functions, then F is a joint distribution function with marginals F_1 and F_2, (Nelsen, 2007).

Farlie–Gumbel–Morgenstern (FGM) copula

Formula for distribution function

$$C(u, v) = uv[1 + \alpha(1 - u)(1 - u)], \quad -1 \leq a \leq 1. \tag{6.5}$$

Formula for density function

$$c(u, v) = 1 + \alpha(1 - 2u)(1 - 2v). \tag{6.6}$$

The correlation coefficient is $\rho = \frac{\alpha}{3}$, which clearly ranges from –3 to 3. After the marginals have been transformed to distributions other than uniform, Gumbel Gumbel and Mustafi (1967) and Schucany Schucany et al. (1978) showed that (i) ρ cannot exceed $\frac{1}{3}$ and (ii) determined it for some well-known distributions, for example, $\frac{\alpha}{\pi}$ for Normal marginals and $\frac{\alpha}{4}$ for exponential ones (Nelsen, 2007).

Archimedean Copula

In some situations, there exists a function φ such that

$$\varphi(C(u, v)) = \varphi(u) + \varphi(v). \tag{6.7}$$

Copulas of the form above are called Archimedean copulas (Genest and MacKay, 1986). Equivalently, we have

$$\varphi(F(x, y)) = \varphi(F_1(x)) + \varphi(F_2(y)); \tag{6.8}$$

i.e., we can write $F(x, y)$ as a sum of functions of marginals F_1 and F_2. Since we are interested in expressions that we can use for the construction of copulas, we want to solve the relation $\varphi(C(u, v)) = \varphi(u) + \varphi(v)$. We thus need to find an appropriately defined inverse $\varphi^{[-1]}$ so that

$$C(u, v) = \varphi^{[-1]}\varphi(u) + \varphi(v) \tag{6.9}$$

Definition

Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(1) = 0$. The pseudoinverse of φ is the function $\varphi^{[-1]}$, with domain $[0,\infty]$ and range $[0,1]$, given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \tag{6.10}$$

Note that if $\varphi(0) = \infty$, then $\varphi^{[-1]}(t) = \varphi^{-1}(t)$ and

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{6.11}$$

C is a copula if and only if the pseudoinverse (or inverse if $\varphi(0) = \infty$) is a convex decreasing function; see Nelsen (2006, p. 111) for a proof, (Nelsen, 2007).

The function φ is called a generator of the copula. If $\varphi(0) = \infty$, we then say that φ is a strict generator and $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ is said to be a strict Archimedean copula. Nelsen (2007) and Kotz et al. (2004) have given several examples of Archimedean copulas.

Gaussian Copula

There are several parametric families of copula functions, such as Student's t copula and Archimedean copulas. One of these families is the Gaussian copula function.

Definition

The copula associated to the joint standard Gaussian distribution is called Gaussian copula. A Gaussian copula is a distribution over the unit cube $[0,1]^d$ constructed from a multivariate Gaussian distribution over $\mathbb{R}$ by using the probability integral transform. For a given correlation matrix $R \in [-1,1]^{d \times d}$, the Gaussian copula with parameter matrix R can be written as

$$C_R^{\text{Gauss}}(u) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right), \tag{6.12}$$

where Φ^{-1} is the inverse CDF of a standard Gaussian and Φ_R is the joint CDF of a multivariate Gaussian distribution with mean vector zero and covariance matrix equal to the correlation matrix R. While there is no simple analytical formula for the copula function, $C_R^{\text{Gauss}}(u)$, it can be upper or lower bounded, and approximated using numerical integration (Botev, 2017). The density can be written as:

$$c_R^{\text{Gauss}}(u) = \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix}^T \cdot \left(R^{-1} - I\right) \cdot \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix} \right), \tag{6.13}$$

where $\mathbf{I}$ is the identity matrix (Arbenz, 2013).

6.2.2 Measures of Dependence Between Two Variables with a Given Copula

Kendall's Tau

Let (x_i, y_i) and (x_j, y_j) be two observations from (X, Y) of continuous random variables. The two pairs (x_i, y_i) and (x_j, y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) \geq 0$ and discordant if $(x_i - x_j)(y_i - y_j) \leq 0$. Kendall's tau is defined as the probability of concordance minus the probability of discordance,

$$\tau = P[(X - X')(Y - Y') \geq 0] - P[(X - X')(Y - Y') \leq 0] \tag{6.14}$$

where (X, Y) is independent of (X, Y) and is distributed as (X, Y). The sample version of Kendall's τ is defined as

$$t = \frac{c - d}{c + d} = \frac{c - d}{n} \tag{6.15}$$

where c denotes the number of concordant pairs and d the number of discordant pairs from a sample of n observations from (X, Y). Just as F can be-expressed as a function of copula C, Kendall's τ can be expressed in terms of the copula [see, for example, Nelsen (2007)] as

$$\tau = 4\int_0^1 \int_0^1 C(u, v)c(u, vdudv - 1 = 4E(C(U, V)) - 1. \tag{6.16}$$

Let C be an Archimedean copula generated by φ. Then, Genest and MacKay (1986) have shown that

$$\tau = 4E(C(U, V)) - 1 = 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt. \tag{6.17}$$

Spearman's Rho

Like Kendall's τ, the population version of the measure of association known as Spearman's rho (denoted by ρ_S) is based on concordance and discordance. Let (X_1, Y_1),(X_2, Y_2), and (X_3, Y_3) be three independent pairs of random variables with a common distribution function H. Then, ρ_S is defined to be proportional to the probability of concordance minus the probability of discordance for the two pairs (X_1, Y_1) and (X_2, Y_3); i.e.,

$$\rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_3) \geq 0] - P[(X_1 - X_2)(Y_1 - Y_3) \leq 0]) \tag{6.18}$$

Eq.(6.18) is really the grade correlation and can be expressed in terms of the copula as

$$\rho_S = 12\int_0^1 \int_0^1 C(u, v)dudv - 3 = 12E(UV) - 3. \tag{6.19}$$

Rewriting the equation above as

$$\rho_S = \frac{E(UV) - \frac{1}{4}}{\frac{1}{12}} \tag{6.20}$$

we simply observe that Spearman's ρ between X and Y is simply Pearson's product-moment correlation coefficient between the uniform variates U and V, (Nelsen, 2007) .

Tests of Independence Against Positive Dependence

Let us consider the problem of testing the null hypothesis of independence,

$$H_0 : F(x, y) = F_1(x)F_2(y), for\ all\ x, y, \tag{6.21}$$

against the alternative of positive quadrant dependence,

$$H_A : F(x, y) \geq F_1(x)F_2(y), for\ all\ x, y, \tag{6.22}$$

with strict inequality holding on a set of nonzero probability. This problem was first considered by Lehmann (1966) who proposed the Kendall's tau and Spearman's correlation tests. Since then, a large number of tests have been proposed in the literature for this hypothesis testing problem; [see, for example, Joag-Dev et al. (1983) and Schriever (1987)].

6.2.3 Maximum Entropy Method joined with Copula

The general form of the maximum entropy problem is to maximize the Shannon's entropy (Thomas and Cover, 2006):

$$h(X) = -\int f(x)logf(x)dx, \tag{6.23}$$

Subject to the moment constraints

$$E(\phi_k(x)) = \int \phi_k(x)f(x)dx = \mu_k, \tag{6.24}$$

where $\mu_0 = 1$, $\phi_0(x) = 1$, $\phi_k(x)$, $k = 0, ..., N$ are $N + 1$ known functions, and $\mu_k, k = 0, ..., N$ are the given expectation data. $\phi_k(x)$, $k = 0, ..., N$ can be in any functional form such as x^n,$\log(x)$, $x\log(x)$ or trigonometric and geometric functions and the main vehicles to determine the required known function is the relationship of the maximum entropy distribution with the Exponential family (Casella and Berger, 2002; Ebrahimi et al., 2008; Pougaza and Djafari, 2011). The moment constraints $\mu_1, ..., \mu_m$, which are normally obtained numerically from data set using the Taylor's theorem (Casella and Berger, 2002). Using an appropriate optimization method, where the objective function is Shannon's entropy, the general form of the maximized density is obtained as follows:

$$f(x) = e^{-\sum_{k=0}^{N} \lambda_k \phi_k(x)}, \qquad x \in S, \tag{6.25}$$

where λ_k should be chosen such that $f(x)$ in Eq.(6.25) satisfies the known moment constraints in Eq.(6.24). Thomas and Cover (Thomas and Cover, 2006) proved that the distribution in Eq.(6.25) has the maximum entropy. The parameters $\lambda = [\lambda_0, ..., \lambda_N]$ should be calculated to find the class of the maximum entropy distributions.

In order to find the bivariate maximum entropy PDF $f(x, y)$, the marginal distributions become the constraints:

$$\begin{cases} \phi_0(x,y) : \int\int f(x,y)dxdy = 1, \\ \phi_1(x,y) : \int f(x,y)dy = f_1(x),\ \forall x, \\ \phi_2(x,y) : \int f(x,y)dx = f_2(y),\ \forall y. \end{cases} \tag{6.26}$$

Hence, the goal is to find the bivariate density $f(x,y)$ compatible with available information in the sense of the maximum entropy principle. Among all possible $f(x,y)$ satisfying the constraints Eq.(6.26), we select the one which optimizes the entropy $h(X,Y)$:

$$h(X,Y) = -\int\int f(x,y)\log f(x,y)dxdy \tag{6.27}$$

via

$$\hat{f} := maximize\ h(X,Y)\ subject\ to\ \phi_k s\ in\ Eq.(6.26)$$

Because the constraints are linear, the choice of a concave objective function $h(X,Y)$ guarantees the existence of a unique solution to the problem. Many entropy functions can serve as concave objective functions, but we focus on the Shannon's entropy (Shannon, 1948a)

$$f(x,y) = exp(-\lambda_1\phi_1(x,y) - \lambda_2\phi_2(x,y) - \lambda_0\phi_0(x,y)) \tag{6.28}$$

where λ_1, λ_2 and λ_0 are obtained by replacing these expressions in the constraints Eq.(6.26) and solving the resulting system of equations. For Shannon's entropy, the constraints can be solved analytically and the joint distribution becomes

$$f(x,y) = f_1(x)f_2(y). \tag{6.29}$$

With the bivariate density obtained from the maximum entropy principle, we can immediately find the corresponding bivariate copula. For the case of Shannon's entropy we have:

$$\begin{aligned} F(x,y) &= \int_0^x\int_0^y f(s,t)dsdt \\ &= \int_0^x\int_0^y f_1(s)f_2(t)dsdt \\ &= \int_0^x f_1(s)ds\int_0^y f_2(t)dt \end{aligned} \tag{6.30}$$

The CDF becomes

$$F(x,y) = F_1(x)F_2(y), \tag{6.31}$$

and the copula is

$$C(u,v) = F(F_1^{-1}(u), F_2^{-1}(v)) = uv. \tag{6.32}$$

In this case, the maximum entropy copula obtained from the Shannon' entropy is the well-known independent copula, which describes independence between two random variables (Pougaza and Djafari, 2011).

6.3 Data Application

6.3.1 Data Description

For further evaluation, we apply our methods to previously analyzed fluorescence microscopy images provided by the Department of Radiation Oncology, University Hospital of Munich. These data have been previously analyzed in Seiler et al. (2011). Here, DNA double strand breaks were marked using $\gamma - H2AX$ (red channel). Using these results of the previous analysis, we can assume co-localization with 53BP1, anti-co-localization with H3K4me3 and independence with H3K9me2 as green channel, see 6.1.

The proposed univariate maximum entropy distribution for the red and green channels is considered as generalized Exponential distribution based on the following constraints

$$\begin{cases} \phi_0(x) : \int f(x)dx = 1, \\ \phi_1(x) : \int xf(x)dx = \mu_1, \\ \phi_2(x) : \int ln(1 - e^{-\lambda x})f(x)dx = \mu_2 \end{cases} \tag{6.33}$$

where μ_1, μ_2 are the known measures based on the data.

Bivariate Generalized Exponential Distribution based on FGM Copula

The univariate generalized exponential distribution has the distribution function

$$F(x; \alpha, \lambda) = (1 - e^{-\lambda x})^{\alpha}; \quad x > 0; \quad \alpha, \lambda > 0 \tag{6.34}$$

and the density function;

$$f(x; \alpha, \lambda) = \alpha\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha-1}; \quad x > 0; \quad \alpha, \lambda > 0 \tag{6.35}$$

where α, λ are the shape and scale parameters respectively. Considering the Eq.(6.34) and Eq.(6.35), the bivariate generalized exponential distribution based on FGM (Eq.(6.5) and Eq.(6.6)) copula can be expressed as

$$\begin{aligned} F(x, y) = {} & (1 - e^{-\lambda_1 x})^{\alpha_1}(1 - e^{-\lambda_2 y})^{\alpha_2} \\ & \times [1 + \theta(1 - (1 - e^{-\lambda_1 x})^{\alpha_1})(1 - (1 - e^{-\lambda_2 y})^{\alpha_2})] \\ & for\ all\ x, y > 0; \quad \alpha, \lambda > 0, \end{aligned} \tag{6.36}$$

and the density function is

$$\begin{aligned} f(x, y) = {} & \alpha_1\alpha_2\lambda_1\lambda_2 e^{-\lambda_1 x}e^{-\lambda_2 y}(1 - e^{-\lambda_1 x})^{\alpha_1}(1 - e^{-\lambda_2 y})^{\alpha_2} \\ & \times [1 + \theta(1 - 2(1 - e^{-\lambda_1 x})^{\alpha_1})(1 - 2(1 - e^{-\lambda_2 y})^{\alpha_2})] \\ & for\ all\ x, y > 0; \quad \alpha, \lambda > 0. \end{aligned} \tag{6.37}$$

Correlation	Localization	MEC(ρ)-Gaus.	MEC(ρ)-FGM	MEC(ρ)-Archmdn	n
Kendal's τ	Co-localization	0.9112	0.2308	0.9007	
Spearman	(53BP1)	0.9017	0.2989	0.8921	

Table 6.2: Estimated Correlation Coefficients and Significance Test Results of $\rho = 0$ & significance level $\alpha = 0.01, 0.05$ for all data.

Correlation	Localization	MEC(ρ)-Gaus.	MEC(ρ)-FGM	MEC(ρ)-Archmdn	n
Kendal's τ	Anti-Co-Localization	0.0227≈ 0	0.0021≈ 0	0.0201 ≈ 0	
Spearman	(H3K4me3)	0.0231≈ 0	0.0021≈ 0	0.0206≈ 0	

Table 6.3: Estimated Correlation Coefficients and Significance Test Results of $\rho = 0$ & significance level $\alpha = 0.01, 0.05$ for all data.

6.3.2 Results

The correlation coefficients for all three methods show a strong positive dependence between $\gamma - H2AX$ and 53BP1, which fits the known co-localization. The negative correlation between $\gamma - H2AX$ and H3K4me3 shows the anti-colocalization. However, as already seen in the simulation study, anti-colocalization results in an absolutely lower value of negative correlation compared to colocalization. Finally, for $\gamma - H2AX$ and H3K9me2 there is no overlap and, hence, correlation is near zero, see Table **??**.

6.4 Discussion & Conclusions

The entropy and copula concepts have been shown to be powerful tools in various applications in hydrology, climatology, and other areas. In recent years, new methods have been proposed for developing copulas based on the maximum entropy theory (entropy–copula). These copulas provide the opportunity to derive probability distribution function of multiple dependent variables and their dependence structure. This study reviews the recent developments in entropy–copula and broadly classifies them into three main groups based on their model structures: 1) continuous maximum entropy copula (CMEC), 2) mixed maximum entropy copula (MMEC), and 3) discrete density maximum entropy copula (DDMEC). The three categories of the entropy– copula models differ in the type of constraints (i.e., uniformly distributed marginal and

Experiment	Localization	MEC(ρ)-Gaus.	MEC(ρ)-FGM	MEC(ρ)-Archmdn	n
Kendal's τ	Independence	-0.3232	-0.2073	-0.3304	
Spearman	(H3K9me2)	-0.3116	-0.2741	0.3156	

Table 6.4: Estimated Correlation Coefficients and Significance Test Results of $\rho = 0$ & significance level $\alpha = 0.01, 0.05$ for all data.

dependence structure) used to derive the copula density within the maximum entropy framework.

.

Chapter 7

General Discussion, Conclusions and Prospective Research Directions

Determining probability density of a random variable based on observations is a major and old issue in statistics. In recent years, various parametric and non-parametric methods have been introduced for determining the class of different statistical distributions. A standard way to estimate an unknown density is to identify its associated properties, such as symmetry, mode, and amplitude, so that choosing a distribution that applies to these properties and estimating its parameters let us approach the main goal. Choosing the best estimator is one challenge for estimating an unknown density. In general, in the case of continuous random variables, a probability density function (PDF) assigns a probability for the observation of a value falling within a specific given range. Empirically determining a PDF corresponding to N samples of univariate data has been investigated extensively in mathematics, with ubiquitous significance for practical applications. Multiple estimation approaches have been used with success for fitting a random data sample to parameters of a known functional form. Nevertheless, the functional form describing the underlying random process is often unknown.

As we have mentioned in the first part of this thesis, the maximum entropy method is used today as a major method for estimating and determining the probability density with high accuracy and efficiency and low bias. This method is employed to obtain the unknown density by solving optimization problems and regarded as one of the most efficient methods capable of yielding maximum possible information for unknown density using the limited and known available information.

The idea is that the principle of the maximum entropy method leads to the selection of a probability density function that is consistent with our knowledge and introduces no unwarranted information. Any probability density function satisfying the constraints that have smaller entropy will contain more information (less uncertainty), and thus says something stronger than what we are assuming. In fact, to the contrary, the principle of maximum entropy guides us to the best probability distribution that reflects our current knowledge and it tells us what to do if experimental data does not agree with predictions coming from our chosen distribution.

We have presented a method allowing estimating the kinetic parameters in DCE-MRI data along with a typically unknown input function. To this end, we use a combination of both maximum entropy and maximum a posterior method. We estimate Eq.(**??**) as a first lead and gain a final estimation of the input function, which allows us to estimate the more reliable kinetic parameters, see Chapter3.

In Chapter4, we have presented the Maximum Entropy Copula (MEC) method for the robust determination of colocalization in microscopy imaging. We proposed to determine an appropriate bivariate distribution from its marginal distributions– which are the separate channels – in order to get their nonlinear correlation using Kendall's τ. The bivariate maximum entropy itself is a powerful method to reconstruct the bivariate distribution using some known constraints. However, with this method, the information about the correlation between two signals will be considered as a constraint to maximize the Shannon entropy. Although, the MEC does not need the correlation of the marginal distributions during computational steps. Hence we prefer the MEC. Interesting enough, the results of applying the MEC and then Kendall's τ showed the correct dependence or independence structure between two signals, though we did not know a priori if there is any correlation.

The results of maximum entropy and maximum a posterior application in Chapter 3 & 5 indicate that the method yields the input function, and the more reliable kinetic parameters that are meaningful from a medical point of view and can lead to new insights into the data. Applications are of course not restricted to the medical area.

Maximum entropy copula is a statistical model class that aims at finding a relationship between scalar and image covariates and a scalar response. We link the concept of dependency in the structure of colocalization in two or more channels to an important problem in statistics which is to estimate a joint probability distribution from its marginals. From a statistical point of view, colocalization and anti-colocalization can be seen as a type of correlation between the two measurements. However, in non-signal areas, which is actually most of the voxels, the two measurements are just independent noise. The actual colocalization or anti-colocalization is only apparent at some voxels, where the signal is high. Hence, one cannot assume a linear relationship between the measurements.

In the next application, we have modified the first algorithm in Chapter3 by adding a new optimization method in the parameter estimation phase, to get more appropriate model fit to data. The results of the simulation study show that the model assumptions can indeed have a quite strong influence on the parameter estimation results and the performance depends on how well the unknown probability function matches the model assumptions, Chapter5.

In general, a concise introduction to entropy maximization and its applicability for deriving models from biological datasets, especially in medical sciences and image processing is provided.

.

Bibliography

Abramov, R. (2006). A practical computational framework for the multidimensional moment-constrained maximum entropy principle. *Journal of Computational Physics*, 211(1):198–209.

Abramov, R. V. (2007). An improved algorithm for the multidimensional moment-constrained maximum entropy problem. *Journal of Computational Physics*, 226(1):621–644.

Abramov, R. V. (2009). The multidimensional moment-constrained maximum entropy problem: A bfgs algorithm with constraint scaling. *Journal of Computational Physics*, 228(1):96–108.

Abramov, R. V. et al. (2010). The multidimensional maximum entropy moment problem: A review of numerical methods. *Communications in Mathematical Sciences*, 8(2):377–392.

Abramov, R. V. and Majda, A. J. (2004). Quantifying uncertainty for non-gaussian ensembles in complex systems. *SIAM Journal on Scientific Computing*, 26(2):411–447.

Adler, J. and Parmryd, I. (2010). Quantifying colocalization by correlation: the pearson correlation coefficient is superior to the mander's overlap coefficient. *Cytometry Part A*, 77(8):733–742.

Aggarwal, V. and Gupta, A. (2019). Integrating morphological edge detection and mutual information for nonrigid registration of medical images. *Current Medical Imaging*, 15(3):292–300.

Aghakouchak, A. (2014). Entropy–copula in hydrology and climatology. *Journal of Hydrometeorology*, 15(6):2176–2189.

Akdağ, S. A. and Dinler, A. (2009). A new method to estimate weibull parameters for wind energy applications. *Energy conversion and management*, 50(7):1761–1766.

Almaraz-Damian, J.-A., Ponomaryov, V., Sadovnychiy, S., and Castillejos-Fernandez, H. (2020). Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. *Entropy*, 22(4):484.

Arbenz, P. (2013). Bayesian copulae distributions, with application to operational risk management–some comments. *Methodology and Computing in Applied Probability*, 15(1):105.

Arsenin, V. J. and Timonov, A. (1983). Building on the basis of local regularization of algorithms of finding approximated resolutions integral equations of the first kind of the convolution type with application of additional information. *Prepr. Keldysh Inst. Appl. Math*, 40:25.

Bain, L. J. and Antle, C. E. (1967). Estimation of parameters in the weibdl distribution. *Technometrics*, 9(4):621–627.

Balakrishnan, N. and Lai, C. D. (2009). *Continuous bivariate distributions*. Springer Science & Business Media.

Barbaresco, F. and Chevalier, P. (2008). Noncircularity exploitation in signal processing overview and application to radar. IET.

Bender, R. and Heinemann, L. (1995). Fitting nonlinear regression models with correlated errors to individual pharmacodynamic data using sas software. *Journal of pharmacokinetics and biopharmaceutics*, 23(1):87–100.

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer.

Berg, B., Stucht, D., Janiga, G., Beuing, O., Speck, O., and Thovenin, D. (2014). Cerebral Blood Flow in a Healthy Circle of Willis and Two Intracranial Aneurysms: Computational Fluid Dynamics Versus Four-Dimensional Phase-Contrast Magnetic Resonance Imaging. *ASME J Biomech Eng*, 15:041003 doi: 10.1115/1.4026108.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645.

Boltzmann, L. (1871). Analytischer beweis des zweiten hauptsatzes der mechanischen wärmetheorie aus den sätzen über das gleichgewicht der lebendigen kraft. *Wiener Berichte*, 63:712–732.

Boomsma, W., Ferkinghoff-Borg, J., and Lindorff-Larsen, K. (2014). Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol*, 10(2):e1003406.

Borwein, J. (1991). As lewis on the convergence of moment problems. *American Mathematical Society*, 23(1).

Borwein, J. M., Lewis, A. S., and Nussbaum, R. D. (1994). Entropy minimization, dad problems, and doubly stochastic kernels. *Journal of Functional Analysis*, 123(2):264–307.

Botev, Z. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):125–148.

Brix, G., Kiessling, F., Lucht, R., Darai, S., Wasser, K., Delorme, S., and Griebe, J. (2004). Microcirculation and microvasculature in breast tumors: pharmacokinetic analysis of dynamic MR image series. *Magnetic Resonance in Medicine*, 52(2):420–429.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms: 2. the new algorithm. *IMA journal of applied mathematics*, 6(3):222–231.

Bryan, G. H. (1894). The laws of distribution of energy and their limitations. *Report of the British Association for the Advancement of Science*, 64:64–101.

Burg, J. (1975). Maximum entropy spectral analysis [ph. d. thesis].

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208.

Casella, G. and Berger, R. (2002). *statistical inference 2*. Duxbury, CA, USA.

Cavalli, A., Camilloni, C., and Vendruscolo, M. (2013). Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of chemical physics*, 138(9):03B603.

Cesari, A., Reißer, S., and Bussi, G. (2018). Using the maximum entropy principle to combine simulations and solution experiments. *Computation*, 6(1):15.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Cheng, H.-L. M. (2007). T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. *Journal of magnetic resonance imaging : JMRI*, 25(5):1073–8.

Cheng, H.-L. M. (2008). Investigation and optimization of parameter accuracy in dynamic contrast-enhanced MRI. *Journal of Magnetic Resonance Imaging*, 28(3):736–743.

Chow, G. C. and Lin, A.-L. (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association*, 71(355):719–721.

Cocco, S., Leibler, S., and Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proceedings of the National Academy of Sciences*, 106(33):14058–14062.

Cofré, R., Herzog, R., Corcoran, D., and Rosas, F. E. (2019). A comparison of the maximum entropy principle across biological spatial scales. *Entropy*, 21(10):1009.

Conrad, K. (2004). Probability distributions and maximum entropy. *Entropy*, 6(452):10.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *Journal of the London Mathematical Society*, 1(4):290–294.

Csiszár, I. (1997). Information theoretic methods in probability and statistics. In *Proceedings of IEEE International Symposium on Information Theory*, page 2. IEEE.

De Martino, A. and De Martino, D. (2018). An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, 4(4):e00596.

Deebani, W. and Kachouie, N. N. (2018). Ensemble correlation coefficient. In *ISAIM*.

Demandolx, D. and Davoust, J. (1997). Multicolour analysis and local image correlation in confocal microscopy. *Journal of Microscopy*, 185(1):21–36.

Denisova, N. (2004a). Bayesian reconstruction in spect with entropy prior and iterative statistical regularization. *IEEE Transactions on Nuclear Science*, 51(1):136–141.

Denisova, N. (2004b). A maximum a posteriori reconstruction method for plasma tomography. *Plasma Sources Science and Technology*, 13(3):531.

Denisova, N. (2019). Bayesian maximum-a-posteriori approach with global and local regularization to image reconstruction problem in medical emission tomography. *Entropy*, 21(11):1108.

Dhadialla, P. S., Ohiorhenuan, I. E., Cohen, A., and Strickland, S. (2009). Maximum-entropy network analysis reveals a role for tumor necrosis factor in peripheral nerve development and function. *Proceedings of the National Academy of Sciences*, 106(30):12494–12499.

Dikaios, N., Arridge, S., Hamy, V., Punwani, S., and Atkinson, D. (2014). Direct parametric reconstruction from undersampled (k, t)-space data in dynamic contrast enhanced MRI. *Medical Image Analysis*, 18(7):989–1001.

Djafari, A. M. and Demoment, G. (1990). Estimating Priors In Maximum Entropy Image Processing. *Proceedings of IEEE ICASSP*, pages 2069–2072.

DJAFARI, M. (1994). 1-maximum d'entropie et problèmes inverses en imagerie. *Traitement du signal*.

Dunn, K. W., Kamocka, M. M., and McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, 300(4):C723–C742.

D'haeseleer, P., Liang, S., and Somogyi, R. (2000). Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726.

Ebrahimi, N., Soofi, E. S., and Soyer, R. (2008.). Multivariate maximum entropy identification, transformation, and dependence. *Multi. Analys*, 99:1217–1231.

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., and Aurell, E. (2013). Improved contact prediction in proteins: using pseudolikelihoods to infer potts models. *Physical Review E*, 87(1):012707.

Elfving, T. (1989.). An Algorithm for Maximum Entropy Image Reconstruction form Noisy Data. *Mathl.comput.Modeling*, 12:729–745.

Fariselli, P., Taccioli, C., Pagani, L., and Maritan, A. (2020). Dna sequence symmetries from randomness: the origin of the chargaff's second parity rule. *Briefings in Bioinformatics*.

Farmer, J. and Jacobs, D. (2018). High throughput nonparametric probability density estimation. *PloS one*, 13(5):e0196937.

Farré, P. and Emberly, E. (2018). A maximum-entropy model for predicting chromatin contacts. *PLoS computational biology*, 14(2):e1005956.

Farsani, Z. A. and Schmid, V. J. (2017). Maximum entropy approach in dynamic contrast-enhanced magnetic resonance imaging. *Methods of information in medicine*, 56(06):461–468.

Farsani, Z. A. and Schmid, V. J. (2020). Co-localization analysis in fluorescence microscopy via maximum entropy copula. *The International Journal of Biostatistics*, 1(ahead-of-print).

Farsani, Z. A. and Schmid, V. J. (2021). Co-localization analysis in fluorescence microscopy via maximum entropy copula. *The International Journal of Biostatistics*, 17(1):165–175.

Fennessy, F. M., McKay, R. R., Beard, C. J., Taplin, M.-E., and Tempany, C. M. (2014). Dynamic contrast-enhanced magnetic resonance imaging in prostate cancer clinical trials: potential roles and possible pitfalls. *Translational oncology*, 7(1):120–129.

Fernandez-de Cossio-Diaz, J. and Mulet, R. (2019). Maximum entropy and population heterogeneity in continuous cell cultures. *PLoS computational biology*, 15(2):e1006823.

Ferrari, U., Obuchi, T., and Mora, T. (2017). Random versus maximum entropy models of neural population activity. *Physical Review E*, 95(4):042321.

Fisher, N. (1997). Encyclopedia of statistical sciences.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368.

Fletcher, P. A., Scriven, D. R., Schulson, M. N., and Moore, E. D. (2010). Multi-image colocalization and its statistical significance. *Biophysical journal*, 99(6):1996–2005.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322.

Frieden, B. R. (1972). Restoring with maximum likelihood and maximum entropy. *JOSA*, 62(4):511–518.

Fritz-Hansen, T., Rostrup, E., Larsson, H. B. W., Sø ndergaard, L., Ring, P., and Henriksen, O. (1996). Measurement of the Arterial Concentration of Gd-DTPA Using MRI: A step toward Quantitative Perfusion Imaging. *Magnetic Resonance in Medicine*, 36(2):225–231.

Frontini, M. and Tagliani, A. (1994). Maximum entropy in the finite stieltjes and hamburger moment problem. *Journal of Mathematical Physics*, 35(12):6748–6756.

Fuglede, B. (1983). The multidimensional moment problem. *Expo. Math.*, 1:47–65.

Ganan, S. and McClure, D. (1985). Bayesian image analysis: An application to single photon emission tomography. *Amer. Statist. Assoc*, pages 12–18.

Garini, Y., Young, I. T., and McNamara, G. (2006). Spectral imaging: principles and applications. *Cytometry Part A*, 69(8):735–747.

Gauthier, M. (2012). Impact of the arterial input function on microvascularization parameter measurements using dynamic contrast-enhanced ultrasonography. *World Journal of Radiology*, 4(7):291.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043.

Golan, A., Judge, G., and Miller, D. (1997). Maximum entropy econometrics: Robust estimation with limited data.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26.

Graeber, T., Heath, J., Skaggs, B., Phelps, M., Remacle, F., and Levine, R. D. (2010). Maximal entropy inference of oncogenicity from phosphorylation signaling. *Proceedings of the National Academy of Sciences*, 107(13):6112–6117.

Granot-Atedgi, E., Tkačik, G., Segev, R., and Schneidman, E. (2013). Stimulus-dependent maximum entropy models of neural population codes. *PLoS Comput Biol*, 9(3):e1002922.

Gull, S. F. and Daniell, G. J. (1978). Image reconstruction from incomplete and noisy data. *Nature*, 272(5655):686–690.

Gumbel, E. J. and Mustafi, C. K. (1967). Some analytical properties of bivariate extremal distributions. *Journal of the American Statistical Association*, 62(318):569–588.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52.

Hadamard, J. (1932). *Le probleme de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*, volume 220. Paris Russian translation).

Haddad, W. M., Hui, Q., Chellaboina, V., and Nersesov, S. G. (2007). Hybrid decentralized maximum entropy control for large-scale dynamical systems. *Nonlinear Analysis: Hybrid Systems*, 1(2):244–263.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Haven, K., Majda, A., and Abramov, R. (2005). Quantifying predictability through information theory: small sample estimation in a non-gaussian framework. *Journal of Computational Physics*, 206(1):334–362.

Havrda, J. and Charvát, F. (1967). Quantification method of classification processes. concept of structural a-entropy. *Kybernetika*, 3(1):30–35.

Haydock, R. and Nex, C. (1984). Comparison of quadrature and termination for estimating the density of states within the recursion method. *Journal of Physics C: Solid State Physics*, 17(27):4783.

Haydock, R. and Nex, C. (1985). A general terminator for the recursion method. *Journal of Physics C: Solid State Physics*, 18(11):2235.

Helmuth, J. A., Paul, G., and Sbalzarini, I. F. (2010). Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC bioinformatics*, 11(1):372.

Hess, S. T., Girirajan, T. P., and Mason, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272.

Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., and Marks, D. S. (2012). Three-dimensional structures of membrane proteins from genomic sequencing. *Cell*, 149(7):1607–1621.

Huang, W., Li, X., Chen, Y., Li, X., Chang, M.-C., Oborski, M. J., Malyarenko, D. I., Muzi, M., Jajamovich, G. H., Fedorov, A., et al. (2014). Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. *Translational oncology*, 7(1):153.

Huang, X.-W. and Emura, T. (2019). Model diagnostic procedures for copula-based markov chain models for statistical process control. *Communications in Statistics-Simulation and Computation*, pages 1–23.

Hussain, S. A. (2009). An introduction to fluorescence resonance energy transfer (fret). *arXiv preprint arXiv:0908.1815*.

Idier, J. (2013). *Bayesian approach to inverse problems*. John Wiley & Sons.

Jackson, A., Constable, C., and Gillet, N. (2007). Maximum entropy regularization of the geomagnetic core field inverse problem. *Geophysical Journal International*, 171(3):995–1004.

Jaynes, E. T. (1957a). Information Theory and Statistical Mechanics. *Physics Reveiw*, 106(4):620–630.

Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *Physical review*, 108(2):171.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge university press.

Jennings, R. C., Belgio, E., and Zucchelli, G. (2020). Does maximal entropy production play a role in the evolution of biological complexity? a biological point of view. *Rendiconti Lincei. Scienze Fisiche e Naturali*, pages 1–10.

Jia, H., Peng, X., Song, W., Oliva, D., Lang, C., and Li, Y. (2019). Masi entropy for satellite color image segmentation using tournament-based lévy multiverse optimization algorithm. *Remote Sensing*, 11(8):942.

Joag-Dev, K., Perlman, M. D., Pitt, L. D., et al. (1983). Association of normal random variables and slepian's inequality. *The Annals of Probability*, 11(2):451–455.

Joe, H. (1993). with given margins. *Journal of multivariate analysis*, 46:262–282.

Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, 22(1):47–64.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Justus, C., Hargraves, W., Mikhail, A., and Graber, D. (1978). Methods for estimating wind speed frequency distributions. *Journal of applied meteorology*, 17(3):350–353.

Kabanikhin, S. I. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, 16(4):317–357.

Kalaidzidis, Y., Kalaidzidis, I., and Zerial, M. (2015). A probabilistic method to quantify the colocalization of markers on intracellular vesicular structures visualized by light microscopy. In *AIP Conference Proceedings*, volume 1641, pages 580–587. American Institute of Physics.

Kendall, M. G. (1948). Rank correlation methods.

Khalifa, F., Soliman, A., El-Baz, A., Abou El-Ghar, M., El-Diasty, T., Gimel'farb, G., Ouseph, R., and Dwyer, A. C. (2014). Models and methods for analyzing dce-mri: A review. *Medical physics*, 41(12):124301.

Kim, J.-M., Jung, Y.-S., Sungur, E. A., Han, K.-H., Park, C., and Sohn, I. (2008). A copula method for modeling directional dependence of genes. *BMC bioinformatics*, 9(1):225.

Kleeman, R. (2002). Measuring dynamical prediction utility using relative entropy. *Journal of the atmospheric sciences*, 59(13):2057–2072.

Kobayashi, H., Ogawa, M., Alford, R., Choyke, P. L., and Urano, Y. (2009). New strategies for fluorescent probe design in medical diagnostic imaging. *Chemical reviews*, 110(5):2620–2640.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2004). *Continuous multivariate distributions, Volume 1: Models and applications*, volume 1. John Wiley & Sons.

Kullback, S. (1968). Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243.

Larsson, H. B. W. and Tofts, P. S. (1992). Measurement of blood-brain barrier permeability using dynamic Gd-DTPA scanning –a comparison of methods. *Magnetic Resonance in Medicine*, 24(1):174–176.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153.

Lesne, A. (2014). Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, 24(3).

Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A., and Fedoroff, N. V. (2006). Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proceedings of the National Academy of Sciences*, 103(50):19033–19038.

Li, T., Griffiths, W., and Chen, J. (2017). Weibull modulus estimated by the non-linear least squares method: a solution to deviation occurring in traditional weibull estimation. *Metallurgical and Materials Transactions A*, 48(11):5516–5528.

Liu, Q., Jiang, Z., and Shi, H. (2019). Maximum entropy image segmentation method based on improved firefly algorithm. In *Journal of Physics: Conference Series*, volume 1213, page 032023. IOP Publishing.

Locasale, J. W. and Wolf-Yadlin, A. (2009). Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data. *PloS one*, 4(8):e6522.

Manders, E., Stap, J., Brakenhoff, G., Van Driel, R., and Aten, J. (1992). Dynamics of three-dimensional replication patterns during the s-phase, analysed by double labelling of dna and confocal microscopy. *Journal of cell science*, 103(3):857–862.

Manders, E., Verbeek, F., and Aten, J. (1993). Measurement of co-localization of objects in dual-colour confocal images. *Journal of microscopy*, 169(3):375–382.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Mead, L. R. and Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417.

Meeuwissen, A. M. H. and Bedford, T. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57(1-4):143–174.

Minerbo, G. (1979). Ment: A maximum entropy algorithm for reconstructing a source from projection data. *Computer Graphics and Image Processing*, 10(1):48–68.

Mohammad-Djafari, A. (1991). A matlap program to calculate the maximum entropy distributions, 221–234 pp.

Mohammad-Djafari, A. (1992). A matlab program to calculate the maximum entropy distributions. In *Maximum entropy and Bayesian methods*, pages 221–233. Springer.

Mohammad-Djafari, A. (1996). A full bayesian approach for inverse problems. In *Maximum entropy and Bayesian methods*, pages 135–144. Springer.

Mohammad-Djafari, A. and Demoment, G. (1990). Estimating priors in maximum entropy image processing. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 2069–2072. IEEE.

Mora, T., Walczak, A. M., Bialek, W., and Callan, C. G. (2010). Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*, 107(12):5405–5410.

Moreno, B. and García-Álvarez, M. T. (2011). Analysing the effect of renewable energy sources on electricity prices in spain. a maximum entropy econometric approach.

Morgan, E. C., Lackner, M., Vogel, R. M., and Baise, L. G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1):15–26.

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, 24(3):69–71.

Murase, K. (2004). Efficient method for calculating kinetic parameters using T1-weighted dynamic contrast-enhanced magnetic resonance imaging. *Magnetic resonance in Medicine*, 51(4):858–862.

Nelsen, R. B. (1999). An introduction to copulas. springer&verla, new york. *Inc. https://doi. org/10.1007/978-1-4757-3076-0*.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nghiem, T.-A., Teleńczuk, B., Marre, O., Destexhe, A., and Ferrari, U. (2018). Maximum entropy models reveal the correlation structure in cortical neural activity during wakefulness and sleep. *bioRxiv*, page 243857.

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466(7306):617–621.

Ormoneit, D. and White, H. (1999). An efficient algorithm to compute maximum entropy densities. *Econometric reviews*, 18(2):127–140.

Orton, M. R., Collins, D. J., Walker-Samuel, S., d'Arcy, J. A., Hawkes, D. J., Atkinson, D., and Leach, M. O. (2007.). Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time. *Phys. Med. Biol.*, 52:2393–2408.

Ouyang, Z.-s., Liao, H., and Yang, X.-q. (2009). Modeling dependence based on mixture copulas and its application in risk management. *Applied Mathematics-A Journal of Chinese Universities*, 24(4):393.

Ozer, H. G. (2008). *Residue associations in protein family alignments*. PhD thesis, The Ohio State University.

Parker, G. J., Roberts, C., Macdonald, A., Buonaccorsi, G. A., Cheung, S., Buckley, D. L., Jackson, A., Watson, Y., Davies, K., and Jayson, G. C. (2006). Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magnetic Resonance in Medicine*, 56(5):993–1000.

Peng, J., Pedersoli, M., and Desrosiers, C. (2020). Mutual information deep regularization for semi-supervised segmentation. In *Medical Imaging with Deep Learning*, pages 601–613. PMLR.

Petrovici, M.-A., Damian, C., and Coltuc, D. (2018). Maximum entropy principle in image restoration. *Advances in Electrical and Computer Engineering*, 18(2):75–83.

Pitera, J. W. and Chodera, J. D. (2012). On the use of experimental observations to bias simulated ensembles. *Journal of chemical theory and computation*, 8(10):3445–3451.

Pougaza, D. B. and Djafari, A. M. (2011). Maximum Entropy Copulas. *AIP Conference Proceeding*, pages 2069–2072.

Pougaza, D.-B., Mohammad-Djafari, A., and Bercher, J.-F. (2010). Link between copula and tomography. *Pattern Recognition Letters*, 31(14):2258–2264.

Pressé, S., Ghosh, K., Lee, J., and Dill, K. A. (2013). Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115.

Quadeer, A. A., McKay, M. R., Barton, J. P., and Louie, R. H. (2020). Mpf–bml: a standalone gui-based package for maximum entropy model inference. *Bioinformatics*, 36(7):2278–2279.

Radon, J. (1917). Über die bestimmung von funktionen durch ihre integralwerte längs dewisse mannigfaltigkeiten breichte sachsische akademie der wissenschaften. *Leipzig, Mathematische-Physikalische Klasse*, 69.

Rao, R. V., Savsani, V. J., and Vakharia, D. (2011). Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. *Computer-Aided Design*, 43(3):303–315.

Rao, R. V., Savsani, V. J., and Vakharia, D. (2012). Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problems. *Information sciences*, 183(1):1–15.

Raphan, M. and Simoncelli, E. P. (2007). Empirical bayes least squares estimation without an explicit prior. *NYU Courant Inst. Tech. Report*.

Remacle, F., Kravchenko-Balasha, N., Levitzki, A., and Levine, R. D. (2010). Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. *Proceedings of the National Academy of Sciences*, 107(22):10324–10329.

Rényi, A. et al. (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.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C. (2011). Detecting novel associations in large data sets. *science*, 334(6062):1518–1524.

Romeo, J. S., Tanaka, N. I., and Pedroso-de Lima, A. C. (2006). Bivariate survival modeling: a bayesian approach based on copulas. *Lifetime Data Analysis*, 12(2):205–222.

Ronneberger, O., Baddeley, D., Scheipl, F., Verveer, P. J., Burkhardt, H., Cremer, C., Fahrmeir, L., Cremer, T., and Joffe, B. (2008). Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. *Chromosome research*, 16(3):523.

Rostami, V., Mana, P. P., Grün, S., and Helias, M. (2017). Bistability, non-ergodicity, and inhibition in pairwise maximum-entropy models. *PLoS computational biology*, 13(10):e1005762.

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput Biol*, 5(5):e1000380.

Roux, B. and Weare, J. (2013). On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of chemical physics*, 138(8):02B616.

Rüschendorf, L. (2013). Copulas, sklar's theorem, and distributional transform. In *Mathematical Risk Analysis*, pages 3–34. Springer.

Sanguinetti, G. et al. (2019). Gene regulatory network inference: an introductory survey. In *Gene Regulatory Networks*, pages 1–23. Springer.

Santolini, M., Mora, T., and Hakim, V. (2014). A general pairwise interaction model provides an accurate description of in vivo transcription factor binding sites. *PloS one*, 9(6):e99015.

Schmid, V. J., Whitcher, B., Padhani, A. R., Taylor, N. J., and Yang, G.-Z. (2006). Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 25(12):1627–36.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012.

Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768.

Schriever, B. F. (1987). Monotonicity of rank statistics in some non–parametric: Testing problems. *Statistica neerlandica*, 41(2):99–109.

Schucany, W. R., Parr, W. C., and Boyer, J. E. (1978). Correlation structure in farlie-gumbel-morgenstern distributions. *Biometrika*, 65(3):650–653.

Seiler, D. M., Rouquette, J., Schmid, V. J., Strickfaden, H., Ottmann, C., Drexler, G. A., Mazurek, B., Greubel, C., Hable, V., Dollinger, G., Cremer, T., and Friedl,

A. A. (2011). Double-strand break-induced transcriptional silencing is associated with loss of tri-methylation at h3k4. *Chromosome Research*, 19(7):883–899.

Seno, F., Trovato, A., Banavar, J. R., and Maritan, A. (2008). Maximum entropy approach for deducing amino acid interactions in proteins. *Physical review letters*, 100(7):078102.

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656.

Shannon, C. E. (1948a). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 623–656.

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

Sharan, R. and Karp, R. M. (2013). Reconstructing boolean models of signaling. *Journal of Computational Biology*, 20(3):249–257.

Shevlyakov, G. and Vilchevski, N. (2002). Robustness in data analysis: criteria and methods, ser. *Modern probability and statistics*.

Shin, Y. S., Remacle, F., Fan, R., Hwang, K., Wei, W., Ahmad, H., Levine, R., and Heath, J. R. (2011). Protein signaling networks from single cell fluctuations and information theory profiling. *Biophysical journal*, 100(10):2378–2386.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., and Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, 26(32):8254–8266.

Skilling, J. (1988). The axioms of maximum entropy. In *Maximum-entropy and Bayesian methods in science and engineering*, pages 173–187. Springer.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Sobhani, F., Xu, C., Murano, E., Pan, L., Rastegar, N., and Kamel, I. R. (2016). Hypovascular liver metastases treated with transarterial chemoembolization: assessment of early response by volumetric contrast-enhanced and diffusion-weighted magnetic resonance imaging. *Translational oncology*, 9(4):287–294.

Soize, C. (2008). Construction of probability distributions in high dimension using the maximum entropy principle: Applications to stochastic processes, random fields and random matrices. *International Journal for Numerical Methods in Engineering*, 76(10):1583–1611.

Spaces, P. M. (1983). North-holland series in probability and applied mathematics.

Sparavigna, A. C. (2019). Entropy in image analysis.

Stein, D. W. (1995). Detection of random signals in gaussian mixture noise. *IEEE Transactions on Information Theory*, 41(6):1788–1801.

Stein, R. R., Marks, D. S., and Sander, C. (2015). Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Comput Biol*, 11(7):e1004182.

Stevens, M. and Smulders, P. (1979). The estimation of the parameters of the weibull wind speed distribution for wind energy utilization purposes. *Wind engineering*, pages 132–145.

Stoyanova, R., Huang, K., Sandler, K., Cho, H., Carlin, S., Zanzonico, P. B., Koutcher, J. A., and Ackerstaff, E. (2012). Mapping tumor hypoxia in vivo using pattern recognition of dynamic contrast-enhanced mri data. *Translational oncology*, 5(6):437.

Strassen, V. et al. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439.

Tagliani, A. (1999). Hausdorff moment problem and maximum entropy: a unified approach. *Applied Mathematics and Computation*, 105(2-3):291–305.

Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J. L., Patel, H., Prieto, A., Petrusca, D., Grivich, M. I., Sher, A., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience*, 28(2):505–518.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Thao, N. X., Ali, M., and Smarandache, F. (2019). An intuitionistic fuzzy clustering algorithm based on a new correlation coefficient with application in medical diagnosis. *Journal of Intelligent & Fuzzy Systems*, 36(1):189–198.

Thomas, A. and Cover, T. M. (2006). *Elements of Information Theory*. John Wiley, New Jersey.

Tihonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, 4:1035–1038.

Tikhonov, A., Arsenin, V. Y., Rubashov, I., and Timonov, A. (1984). The first soviet computer tomograph. *Priroda*, 4:11–21.

Tkačik, G., Marre, O., Mora, T., Amodei, D., Berry II, M. J., and Bialek, W. (2013). The simplest maximum entropy model for collective behavior in a neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03011.

Tkačik, G., Prentice, J. S., Balasubramanian, V., and Schneidman, E. (2010). Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424.

Tofts, P. S. and Kermode, A. G. (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging - 1. Fundamental concepts. *Magnetic Resonance in Medicine*, 17(2):357–367.

Tuchler, M., Singer, A. C., and Koetter, R. (2002). Minimum mean squared error equalization using a priori information. *IEEE Transactions on Signal processing*, 50(3):673–683.

Tumanski, S. (2006). *Principles of electrical measurement*. CRC press.

Turchin, V., Kozlov, V., and Malkevich, M. (1971). The use of mathematical-statistics methods in the solution of incorrectly posed problems. *Soviet Physics Uspekhi*, 13(6):681.

Turchin, V. F. (1967). Solution of the fredholm equation of the first kind in a statistical ensemble of smooth functions. *USSR Computational Mathematics and Mathematical Physics*, 7(6):79–96.

Usuda, K., Iwai, S., Funasaki, A., Sekimura, A., Motono, N., Matoba, M., Doai, M., Yamada, S., Ueda, Y., and Uramoto, H. (2019). Diffusion-weighted magnetic resonance imaging is useful for the response evaluation of chemotherapy and/or radiotherapy to recurrent lesions of lung cancer. *Translational oncology*, 12(5):699–704.

Valeur, B. and Berberan-Santos, M. N. (2012). *Molecular fluorescence: principles and applications*. John Wiley & Sons.

von der Linden, W. (1995). Maximum-entropy data analysis. *Applied Physics A*, 60(2):155–165.

Wang, J. and Zabaras, N. (2004). A bayesian inference approach to the inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, 47(17-18):3927–3941.

Wang, S., Arena, E. T., Eliceiri, K. W., and Yuan, M. (2017). Automated and robust quantification of colocalization in dual-color fluorescence microscopy: A nonparametric statistical approach. *IEEE Transactions on Image Processing*, 27(2):622–636.

Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72.

Weinmann, H.-J., Laniado, M., and Mützel, W. (1984). Pharmokinetics of Gd-DTPA/Dimeglumine after intravenous injection into healthy volunteers. *Physiological Chemistry & Physics & Medical NMR*, 16:167–172.

Wu, Z., Phillips Jr, G. N., Tapia, R., and Zhang, Y. (2001). A fast newton algorithm for entropy maximization in phase determination. *SIAM review*, 43(4):623–642.

Xu, W., Hou, Y., Hung, Y., and Zou, Y. (2010). Comparison of spearman's rho and kendall's tau in normal and contaminated normal models. *arXiv preprint arXiv:1011.2009*.

Yang, C., Karczmar, G., Medved, M., and Stadler, W. (2004). Toward local arterial input functions in dynamic contrast-enhanced MRI. *J Magn Reson Imaging*, 52(5):1110–1117.

Yari, G. and Farsani, Z. A. (2015). Application of the maximum entropy method for determining a sensitive distribution in the renewable energy systems. *Journal of Energy Resources Technology*, 137(4):042006.

Yeh, F.-C., Tang, A., Hobbs, J. P., Hottowy, P., Dabrowski, W., Sher, A., Litke, A., and Beggs, J. M. (2010). Maximum entropy approaches to living neural networks. *Entropy*, 12(1):89–106.

Yeo, G. and Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to rna splicing signals. *Journal of computational biology*, 11(2-3):377–394.

Zhang, B. and Wolynes, P. G. (2015). Topology, structures, and energy landscapes of human chromosomes. *Proceedings of the National Academy of Sciences*, 112(19):6062–6067.

Zhang, D., Wells, M. T., and Peng, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis*, 99(4):577–588.

Zhao, S., Wang, Y., Yang, Z., and Cai, D. (2019). Region mutual information loss for semantic segmentation. In *Advances in Neural Information Processing Systems*, pages 11117–11127.

Eidesstattliche Versicherung

(gemäß §8 Abs. 2 Pkt. 5 der Promotionsordnung vom 12.07.2011)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 5. Mai 2021 _____

Zahra Amini Farsani