The History and New Trends of Measuring Dependence: From Bayes, Galton, and Pearson to the 21st Century

Ali Habibnia, Jonathan Gendron and Sanjana Rayani[*]

Abstract

In this study, our objective is to first present a critical examination of the history of the initial conceptualization of statistical dependence developed by Bayes, Galton, Pearson, and subsequently by other scientists. We discuss the various challenges that scientists faced and overcame while developing the procedures for measuring dependence that are now widely used. Then, We clarify the hidden assumptions and limits of various types of correlation measurements, as well as provide practitioners with improved model-free and model-based alternatives from the existing literature. We also analyze the interpretation of the correlation measurements and several instances in which the application of the correlation approach may be improper. The concepts discussed in this study are supported by examples from literature in the field of economics and finance. The ultimate goal of this study is to highlight that statistical measures and methods can mislead us if we blindly generate them without checking the assumptions and limitations of the methods and without considering the nature of the data in our hands. Comparing the benefits and drawbacks of various measures of dependence will allow us to create a road map for practitioners and researchers on which measures of dependence to use in which circumstances. We aim to continue this research in high dimensional space which will provide a crucial contribution to it's regular use by many practitioners in big data and machine learning.

Keywords: Statistical dependence, Pearson's ρ, nonlinear dependence, multivariate dependence, mutual information.

[*]Department of Economics, Virginia Tech, USA. Email of Principal Investigator: habibnia@vt.edu

We would like to extend our sincerest thanks and appreciation to the Kellogg Center for Philosophy, Politics, and Economics (PPE) at Virginia Tech for their invaluable financial and logistical support throughout this research project.

1 Introduction

As scientists, discovering dependency between variables is an essential part of our efforts to understand the world, identify the causes of diseases, assess climate change, explain business cycles or protect against catastrophic events such as hurricanes or economic collapses. To many applied scientists and policymakers, measuring dependence is synonymous with correlation originated by Galton in 1888. Galton, Charles Darwin's cousin, required a measure of association for his hereditary investigations. Later on, in 1896, Pearson gave a more precise mathematical development and introduced the Pearson product-moment correlation coefficient (PPMCC). However, in the 18th century, over 100 years before Sir Francis Galton and Karl Pearson's preoccupation with co-relation and correlation, various authors such as Thomas Bayes (in his 1763 paper) used verbal definitions of (in)dependence on occasion in their writings. Although we have all been taught that there are several serious weaknesses in Pearson correlation such as the non-Gaussianity, nonlinearity, and robustness issues, many researchers, policymakers, and practitioners still make association claims based on misleading correlations without considering the key assumptions of this measurement. These claims are too often unscrutinized, amplified, and mistakenly used to guide decisions. This can lead to mistakes and avoidable disasters, whether it's an individual, a company, or a government that's making the decision.

Every statistical model includes three core probabilistic assumptions of distribution, independence, and heterogeneity. All three must be accounted for to mitigate confounding factors and to ensure inferences are valid (Spanos, 2019; Wu & Mielniczuk, 2010). For example, if dependence is present, this can confound our results if not properly accounted for. That is the motivation of this research, to best understand dependence so it can be best accounted for in model-free methods, model-based methods, as well as network-based methods. Model-free methods are methods that can be used to measure dependence when there is no underlying model to estimate, while model-based methods are methods that require an underlying model to measure dependence.

Statistical dependence refers to the relationship between two or more variables, where the value of one variable provides information about the values of other variables, they are considered to be dependent. In probability analysis, statistical dependence between random variables can be defined in terms of their joint and marginal probability distributions. Two random variables, X and Y, are considered statistically dependent if their joint probability distribution P(X, Y) is not equal to the product of their marginal probability distributions P(X) and P(Y). Mathematically, this can be expressed as: X and Y are dependent if P(X, Y)$\neq$ P(X)P(Y) for some values of X and Y.

Another way to define dependence in probability analysis is through conditional probabilities. If the conditional probability distribution of X given Y (or vice versa) is not equal to the marginal probability distribution of X, then X and Y are dependent: X and Y are dependent if $P(X|Y) \neq$ P(X) or P(Y—X) $\neq$ P(Y) for some values of X and Y. Here, $P(X|Y)$ is the probability distribution of X given Y, and $P(X|Y)$ is the probability distribution of Y given X.

The distinction between dependent and independent events was first made by French mathematician Pierre-Simon Laplace (1749-1827) in his work on probability theory, particularly in his seminal book "Théorie Analytique des Probabilités" (Analytical Theory of Probability) published in 1812 Laplace (1812). However the foundational framework for understanding the relationship between events and updating our beliefs about the likelihood of events based on new information first provided by Thomas Bayes (1701-1761), an English statistician, philosopher, and Presbyterian minister, best known for his work on probability theory and for formulating Bayes' theorem.

Mutual information provides an alternative definition of statistical dependence. Mutual information is a measure of the amount of information that one random variable contains about another random variable. It is based on the concepts of entropy from information theory, and it quantifies the reduction in uncertainty about one variable when the value of the other variable is known. Mutual information as a measure of dependence between random variables was first introduced by Claude E. Shannon in his groundbreaking paper "A Mathematical Theory of Communication," published in 1948. Shannon's work laid the foundation for information theory, a field that studies the quantification, storage, and communication of information Shannon (1948).

For two discrete random variables X and Y, the mutual information I(X; Y) is defined as:

$$I(X;Y) = \sum_{x} \sum_{y} P(x,y) \cdot \log_2 \left(\frac{P(x,y)}{P(x) \cdot P(y)} \right) \tag{1}$$

where I(X; Y) is the mutual information between X and Y, P(x, y) is the joint probability mass function of X and Y, P(x) and P(y) are the marginal probability mass functions of X and Y, respectively, The summations are taken over all possible values of x and y. For continuous random variables, the summations are replaced with integrals, and the probability mass functions are replaced with probability density functions. If the mutual information is greater than 0, it indicates that X and Y are statistically dependent, as knowing the value of one variable reduces the uncertainty about the other variable. It is important to note that mutual information is a more general measure of dependence than the Pearson correlation coefficient, as it can capture non-linear dependencies between variables, while Pearson correlation is limited to linear relationships.

In the context of the frequency domain, statistical dependence between two time series or signals can be measured by evaluating the coherence between their spectral (Fourier) representations. The concept of coherence is used to quantify the linear relationship between signals as a function of frequency. Given two time series or signals, $X(t)$ and $Y(t)$, their respective Fourier transforms are $X(f)$ and $Y(f)$, where f denotes frequency. The cross-spectral density (CSD) between X and Y is defined as:

$$G_{xy}(f) = E[X(f)Y^*(f)] \tag{2}$$

where $E[\cdot]$ is the expectation operator, and $Y^*(f)$ is the complex conjugate of $Y(f)$. The power spectral density (PSD) of X and Y are defined as:

$$G_{xx}(f) = E[X(f)X^*(f)] \tag{3}$$

$$G_{yy}(f) = E[Y(f)Y^*(f)] \tag{4}$$

The magnitude-squared coherence (MSC) between X and Y is a frequency-domain measure of their dependence, defined as:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f) \cdot G_{yy}(f)} \tag{5}$$

where $C_{xy}(f)$ is the MSC at frequency f, and $|\cdot|$ denotes the absolute value. The MSC ranges from 0 to 1, where 0 indicates no linear relationship between the signals at that frequency, and 1 indicates a perfect linear relationship.

Magnitude-squared coherence (MSC) was first proposed by Norbert Wiener and Paul C. Schreiber in their joint work on spectral analysis and time series analysis. In 1962, Paul C. Schreiber published a paper titled "Cross-Spectral Analysis: An Introduction," Schreiber (1962) which introduced the concept of MSC as a measure of the linear relationship between two signals in the frequency domain. Schreiber's work was inspired by and built upon Norbert Wiener's earlier work on time series analysis and spectral theory. The development of MSC and its use as a measure of statistical dependence in the frequency domain has since become an important tool in signal processing, time series analysis, and systems engineering, helping researchers and engineers understand the frequency-specific relationships between signals and processes.

Each of the subsections contained in Section 1 is associated with one of the aforementioned techniques, namely model-free, model-based, and network-based. differences between the measures are analyzed. The paper presents a comprehensive examination of various methods and measures of dependence, including their respective strengths and weaknesses. Additionally, the similarities and distinctions between these measures are thoroughly evaluated. The disparities between each metric are evaluated. The second section provides a comprehensive analysis of the existing literature.

1.1 Brief History

To many applied scientists and policymakers, measuring dependence is synonymous with correlation originated by Galton (1889). Galton, Charles Darwin's half-cousin, required a measure of association for his hereditary investigations. A collection of letters and correspondence between Charles Darwin and Francis Galton can be found on the dedicated website, providing valuable insights into their relationship and collaboration Darwin and Galton (n.d.). Fancher's article Fancher (2009) delves into the personal and intellectual mutual relationship between two thinkers and the mutual influence on each other's work, highlighting the areas of agreement and disagreement between them. The correspondence between Darwin and Galton, which spanned several decades, offers insights into their shared interests, scientific debates, and collaboration on various projects. Although Galton's work on eugenics and his views on selective breeding was met with some skepticism by Darwin, their relationship remained largely amicable and productive.

Later on, in 1896, Pearson Pearson (1896) gave a more precise mathematical development and introduced the Pearson product-moment correlation coefficient (PPMCC). Sir Francis Galton and Karl Pearson had a close professional relationship, with Pearson being one of Galton's most prominent followers and a leading proponent of his ideas. Pearson was a British mathematician and biostatistician who made significant contributions to the development of modern statistics, including the Pearson correlation coefficient, chi-square test, and principal component analysis.mGalton's work on heredity, eugenics, and the study of human intelligence inspired Pearson to apply mathematical methods to the study of biological and social phenomena. As a result, Pearson played a pivotal role in establishing the field of biometrics, which uses statistical techniques to analyze biological data.

In 1904, Pearson was appointed the first Galton Professor of Eugenics at University College London, a position funded by Galton himself. Pearson later founded the Biometrika journal in 1901 and the Annals of Eugenics in 1925, both of which were instrumental in promoting Galton's ideas and advancing the field of biostatistics. Their relationship was not only professional but also personal. Pearson had great admiration for Galton, and their correspondence reflects their mutual respect and collaboration on various projects. After Galton's death in 1911, Pearson took it upon himself to write Galton's biography, titled "The Life, Letters, and Labours of Francis Galton," published in three volumes Pearson (1914-1930).

However, in the 18th century over 100 years before Sir Francis Galton and Karl Pearson's preoccupation with co-relation and correlation, various authors such as Thomas Bayes, in his 1763 paper Bayes (1763), used verbal definitions of (in)dependence on occasion in their writings. This subsection will highlight and review key historical figures to show how statistics and probability theory started, integrated with each other, and evolved over time.

Gerolamo Cardano 1501 - 1576: Gerolamo Cardano to many is considered the first key player of statistics since this polymath formally defined independence as well as the concept of identical distributed trials, and relative frequencies (Cardano, 2015). All of this was shown by with his studies of gambling with dice rolls in his book "The Book on Dice Games", this was published after his death in 1663.

Abraham de Moivre in 1718: In his book "The Doctrine of Chances," this french mathematician Abraham de Moivre was the first to formalize independence and conditional probabilities to create a mathematical quantification for uncertainty. Finding the probability of games of chance through combinatorial methods using binomial coefficients is a taxing process when performing calculations for large values, so there was a need to determine a method that could calculate simple approximations to the binomial coefficient and the binomial distribution. The largest issue was to find a value of n, the number of trials needed to obtain a specified probability. The work of de Moivre was heavily based on deductive reasoning as he introduced the concept of the normal distribution. He observed how having a large n value affected the symmetric binomial $(1 + 1)n$ and found a large sample approximation. In 1721, de Moivre began investigating the nature of the binomial distribution when p was equal to 0.5. After finding an approximation of the maximum term, he found an approximation of the ratio of the maximum term at a distance of d from the maximum. From 1725, James Stirling and de Moivre both worked to find a binomial approximation and landed a constant equal to $\sqrt{2\pi}$. However, they later decided that creating an approximation to $ln(n!)$ would create a simpler form. In 1733, de Moivre simplified his results and demonstrated how the normal distribution could be used to approximate the binomial distribution. A major shortcoming of de Moivre's approach was that he did not account for inverse probability since he focused on his approximation theorem (Hald, 1988).

Thomas Bayes in 1763: In 1763, Thomas Bayes published a paper titled "An Essay towards solving a Problem in the Doctrine of Chances" which presented Bayes theorem. This theorem provides a way to alter the probability of a hypothesis when provided with new evidence. Bayes theorem states that the probability of a hypothesis given some observed evidence is proportional to the product of the prior probability of the hypothesis and the probability of the evidence given the hypothesis, divided by the probability of the evidence. Bayes theorem is a fundamental concept in probability theory that describes how to update beliefs in the face of new evidence. It provides a way to calculate the probability of a hypothesis (or event) given some observed evidence. In other words, it allows us to update our prior beliefs about a hypothesis based on new data or evidence.

Adrien-Marie Legendre in 1805: In 1805, Adrien Marie Legendre was the first to formally introduce the method of least squares in the appendix of his book, "Nouvelles méthodes pour la détermination des orbites des comètes" ("New Methods for Determining the Orbits of Comets") (Stigler, 1986). Legendre came across empirical problems in the fields of astronomy and geodesy. His journey started in 1792 when Legendre was tasked with measuring the length of the meridian quadrant, the distance from the equator to the north pole. During this period, the National Convention initiated a project that would create a new measurement system known as the metric system. The basis of the system was to be a meter, defined as $1/10,000,000$ of the meridian quadrant. Legendre suggested that the measurement error of the multiple angular measures being converted to arc lengths will need to be reduced. In most scientific investigations, the possible results are present in an equation in a form that consists of known coefficients, unknown variables, and a condition of error. In order to determine the value of the unknown variables the condition of error would need to be reduced to nearly zero. However, when the amount of equations present was greater than the amount of unknown variables, it was impossible to determine values of the unknown variables that would eliminate the amount of errors in the equations utilized. Legendre wanted to determine an easy to apply method that would reduce the sum of the squares of errors to the minimum possible value. Legendre's principle was a set of simple linear equations that would need to be solved to determine the minimum amount of error. The equations were derived from differentiating the sum of squared errors.

Carl Friedrich Gauss in 1809: In his 1809 publication, "Theoria Motus Corporum Coelestium in Sectionibus Conicis Solum" ("The Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections"), Carl Friedrich Gauss used his investigation of the mathematics of planetary orbits to introduce the method of least squares (Stigler, 1986). At this point in time, there was no formal connection between probability theory and the least squares method. It was important to find a connection in order to determine how useful and accurate the method could be. Gauss was interested in finding a method to estimate the values of unknown parameters in a system of equations when provided with a set of measurements and a degree of error. Gauss noted that the most probable system of values could be calculated by setting the derivatives of the equation equal to zero and solving the resulting equations. However, these equations had an error curve and Gauss realized that he needed a formula for the error curve. He stated that the arithmetic mean is most probably only when the curve is normally distributed connecting his method to the central limit theorem. Gauss founded the tradition known as the theory of errors, which linked probability theory to the modeling of observed data by operationalizing the Central Limit Theorem. Gauss was finding measurements for the heights of a series of points of the Earth's surface. He recognized that his measurements would have some degree of error due to factors such as imprecise measurement tools and environmental factors. Additionally, he understood that some of his measurements would prove to be more accurate than others, which delineated the theory of errors. To account for this variability in the accuracy of measurements, Gauss developed a method for minimizing the sum of the squares of the differences between each measurement and the estimated value. Gauss linked the method to probability and provided algorithms for the computation of the estimate.

Pierre-Simon Laplace in 1812: The French mathematician and astronomer Pierre-Simon Laplace improved and built on the work of Gauss in his book "The Analytical Theory of Probability" (Stigler, 1986). Laplace played a significant role in further development of the theory of errors. Laplace was particularly interested in the problem of estimating the parameters of a probability distribution from a set of observations. He recognized that the method of least squares developed by Gauss could be extended to apply for any probability distribution. Laplace saw that by a slight extension of Gauss's Central limit theorem, he could show that the error was approximately normal for limiting distribution. Laplace concluded that the least squares method was the most accurate for large numbers of equations because it would provide a method for the smallest expected error and most likely to be near the quantity being estimated.

Pafnuty Chebychev in 1899: Pafnuty Chebyshev and his students, Andrey Markov and Aleksandr Lyapunov, made significant contributions to probability theory in the late 19th century (including dependence) (Stigler, 1986). He is particularly known for his work in "Works by P.L. Chebychev" on the creation of random variables, the weak law of large numbers, the Chebyshev inequality, and the development of Chebyshev polynomials. Chebyshev utilized his understanding of mathematical analysis and applied it to various problems in probability theory. He was particularly interested in problems related to the distribution of errors in measurements and the behavior of random variables. He worked extensively on developing techniques for estimating the probability of events that involve random variables and for analyzing the behavior of random variables in large samples. Chebyshev was one of the first mathematicians to rigorously define what we now call a random variable. He showed that any function of a random variable can also be a random variable, and that the expected value and variance of a random variable can be used to describe its behavior. The weak law of large numbers is a fundamental result in probability theory that states that the sample mean of a sequence of independent and identically distributed random variables converges in probability to the expected value of the random variable.

Sir Francis Galton in 1889: Sir Francis Galton was an English mathematician whose work focused on the relation between two variables which helped lay the foundation for Pearson's correlation coefficient in his work "Co-relations and their measurement, chiefly from anthropometric data" (Stigler, 1986). He developed a technique called regression analysis, which involves fitting a line to a set of data points to describe the relationship between two variables. Galton's regression line was based on the idea of minimizing the sum of the squared vertical distances between the data points and the line. Galton also developed the concept of correlation, which measures the degree to which two variables are related to each other. He introduced the idea of the correlation coefficient, which is a measure of the strength and direction of the relationship between two variables. Galton's correlation coefficient was based on the idea of measuring the angle between the two regression lines that describe the relationship between the two variables. However, Galton's correlation coefficient had some limitations, as it was not always easy to interpret and was sometimes unstable when applied to small datasets.

Karl Pearson in 1905: Karl Pearson was a British mathematician known for his work on the theory of correlation and the development of the Pearson correlation coefficient in "Contributions to the mathematical theory of evolution XIV. On the general theory of skew correlation and non-linear regression" (Stigler, 1986). Pearson's development of this coefficient was based on his work in biometrics, where he was interested in the relationship between physical traits such as height, weight, and head circumference. He recognized that there was a need for a quantitative measure of the strength and direction of the relationship between two variables, and the Pearson correlation coefficient provided a solution to this problem. Pearson's correlation coefficient is a measure of the linear relationship between two variables. It is defined as the covariance of the two variables divided by the product of their standard deviations. The coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. In addition to his work on correlation, Pearson made many other important contributions to statistics, including the development of the χ^2 test, which is used to test the independence of two categorical variables, and the method of moments, which is used to estimate the parameters of a statistical model.

George Udny Yule in 1912: George Udny Yule, a student of Karl Pearson, introduced dependence measures that addressed Pearson's limitations in "An introduction to the theory of statistics" (Stigler, 1986). In his 1907 book, "An Introduction to the Theory of Statistics," Yule recognized that Pearson's correlation coefficient had limitations in its ability to capture non-linear relationships between variables and was also sensitive to outliers. To address these issues, Yule developed two alternative measures of association, known as Yule's Q and Yule's Y. Yule's Q is a measure of association that is particularly well-suited for binary data, such as the presence or absence of a particular characteristic. It measures the difference between the proportions of one group that has the characteristic and the proportion of the other group that has it. Yule's Y, on the other hand, is a measure of association that is more general and can be used with any type of data. It measures the odds ratio of one variable given the value of the other variable.

R.A. Fisher in 1922: Ronald A. Fisher was a British statistician and geneticist who is considered the father of modern frequentist statistics. In his 1922 paper, "On the Mathematical Foundations of Theoretical Statistics," he established the foundations of statistical hypothesis testing (Fisher, 1922). Fisher also introduced the concept of null and alternative hypotheses, and developed the notion of the p-value, which is a measure of the evidence against the null hypothesis. The concept of the level of significance was also introduced, this is a measure of how confident we are that the null hypothesis is false. Fisher's work on hypothesis testing was groundbreaking and had a profound impact on the field of statistics. It established a rigorous framework for testing scientific hypotheses and has become the foundation of modern statistical inference. His approach, which emphasizes the importance of collecting data and testing hypotheses rigorously, is still widely used today in scientific research.

Jerzy Neyman & Egon Pearson in 1933: Jerzy Neyman and Egon Pearson created hypothesis testing in, "On the Problem of the Most Efficient Tests of Statistical Hypotheses" (Neyman & Pearson, 1933). They proposed a framework for testing statistical hypotheses based on the notion of statistical significance, which they defined as the probability of observing data at least as extreme as the observed data, assuming that the null hypothesis is true. Neyman and Pearson introduced the concept of a statistical test, which is a procedure for determining whether the observed data provide enough evidence to reject the null hypothesis. They also introduced the concept of the level of significance and proposed that a statistical test should be designed to maximize the power of the test. The Neyman and Pearson approach to hypothesis testing emphasizes the importance of power and the control of type I errors, has become the foundation of modern statistical inference. Their work laid the groundwork for the development of many other important statistical concepts, including the concept of the p-value, which is a measure of the evidence against the null hypothesis, and the concept of confidence intervals, which provide a range of plausible values for the parameter of interest.

Andrey Kolmogorov in 1933: Andrey Kolmogorov provided key math foundations of probability theory, published in, "Foundations of Probability Theory" (Kolmogorov, 1933). Kolmogorov introduced the concept of a probability space, which is a mathematical model that describes the possible outcomes of a random experiment. A probability space consists of a set of possible outcomes, a set of events, and a probability measure that assigns probabilities to each event. This mathematical model allowed probability theory to be studied and developed in a rigorous and systematic way. Kolmogorov also introduced the concept of conditional probability, which is the probability of an event given that another event has occurred. He developed the axioms of probability theory, which are a set of mathematical rules that govern the behavior of probabilities. These axioms provide a foundation for probability theory that is based on rigorous mathematical principles.

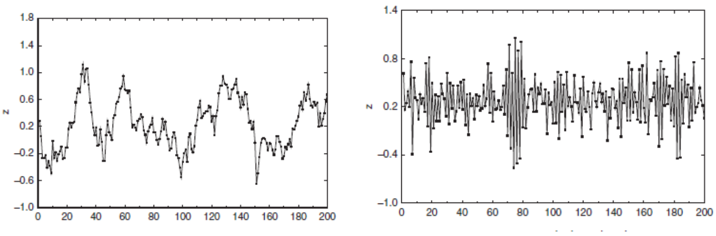
1.2 Model Free Methods

1.2.1 Definitions

Independence: Long before Sir Francis Galton and Karl Pearson discussed independence in the 19th century, in 1663 the polymath Gerolamo Cardano formally defined independence with events A & B in the same domain as follows (Cardano, 2015).

$$\mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B) \tag{6}$$

Dependence: If any part of the above equation fails, it can be said that independence is no longer exhibited, but dependence is exhibited. Earlier dependence was observed if a variable X_t changes because of changes in another variable Y_t, but how exactly does it change? Dependence can occur in two ways: as positive or negative dependence (Spanos, 2019). Positive dependence means that X_t will increase due to an increase in Y_t and vice-versa, while negative dependence means that X_t will decrease due to an increase in Y_t and vice-versa. The figure below on the left illustrates positive dependence, while the figure below on the right illustrates negative dependence (Spanos, 2019):



(a) Positive dependence (b) Negative dependence

Figure 1: Both graphs are normal and identically distributed, their difference is in the type of dependence they exhibit

Dependence can also be observed between time periods: for instance, a variable X_t changes because of changes in it's prior observations X_{t-1} (Spanos, 2019).

Dependence & Bayes' Theorem: Regarding dependence above, there is a theorem called Bayes' theorem. This theorem measures the conditional probability of event A_i depending on event B (Spanos, 2019). In other words, if A_i is dependent on B, this conditional probability measure should be used to account for this dependence.

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i) \cdot \mathbb{P}(B|A_i)}{\sum_{i=1}^{n} \mathbb{P}(A_i) \cdot \mathbb{P}(B|A_i)}, \mathbb{P}(B) > 0 \quad (7)$$

Linearity: A linear relationship captures a constant slope between at least two variables, and can typically be seen as a straight line (at whatever angle) (Wooldridge, 2012). For instance, if X_t & Y_t are linearly related, their relationship can be shown as $Y_t = m \cdot X_t$. Whether X_t is 5 or 100, it'll still be multiplied by the same m. That describes linearity in data, which essentially observes the relationship between two (or more) variables. But there is another type of linearity called linearity in parameters (Spanos, 2019). Instead of just comparing the two (or more) variables as they are, they are compared while within a statistical model. Formally speaking, parameters of a statistical model exhibit linearity when changes in an independent variable (typically shown as X_t) are multiplied by it's constant parameter (typically shown as β) to obtain the estimated value of the dependent variable (typically shown as Y_t). The above description does not include an error term but is excluded only for the sake of the explanation, it should be included in a proper statistical model. Linearity is often a requirement to use certain measures of dependence such as correlations and regressions, linearity's full relation to dependence will be further discussed the following Limitations subsection.

Comovement: Variables move together, either in the same or opposite direction. For example when X_t increases, Y_t decreases (Spanos, 2019). This term includes both positive and negative dependence.

Covariance: One of the most basic measures of dependence, this looks at the variance between two variables (Delicado & Smrekar, 2009; Spanos, 2019; Tjøstheim, Otneim, & Støve, 2018).

$$Cov(X,Y) = E(X - E(X))(Y - E(Y)) \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{8}$$

This can also be represented in the following way.

$$Cov(X,Y) = \frac{1}{N}\sum_{i}^{N}(x_i - \bar{x})(y_i - \bar{y}) \tag{9}$$

Note that covariance is typically seen as a n x n matrix with a diagonal of 1, where n is the number of variables.

Correlation: Also referred to as Pearson's Correlation Coefficient, this is the most commonly used measure of dependence, taking values between -1 & +1. The higher the absolute value is, the more dependent each variable is on the other (Spanos, 2019; Tjøstheim et al., 2018). It is important to note that there are many limitations that are not accounted for in a variety of fields, this will be discussed further in the Limitations subsection.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}, Var(X) = E(X^2) - [E(X)]^2 \tag{10}$$

This can also be represented in the following way.

$$Corr(X,Y) = \frac{1}{N}\frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2}} \tag{11}$$

A correlation heatmap is a very helpful visual tool that can highlight which relationships are and are not dependent (Uddin, Rahman, Hedström, & Ahmed, 2019). This is achieved by looking at variables pairwise; for instance the correlation of CRIM is measured for every single variable one at a time (including itself). Note that this heatmap can also be made for covariance too.

Copula: It is argued that copulas can model dependency for high dimensions, this is based on Sklar's 1959 theorem. More precisely, uniform CDFs ($F_i(x_i)$, $i = 1, ..., p$) can be transformed into a joint CDF $F(x_1, ..., x_p)$ (Genest & Neslehova, 2014; Rüschendorf, 2009; Sklar, 1959; Spanos, 2019; Tjøstheim et al., 2018).

$$F(x_1, ..., x_p) = C(F_1(x_1), ..., F_p(x_p)) \tag{12}$$

High Dimensions: This refers to modeling data where the number of variables are higher than the sample size (Buhlmann, 2017). Dependence cannot be modeled in the same way for high dimensional data.

1.2.2 Limitations

The prior subsection introduced the following common measures of dependence: covariance, correlation, and copulas. Each of these measures have their own limitations that need to be accounted for if they are to be used. For instance, linearity is required in order to use covariance or correlation. This part not only discusses limitations, but also shows how newer measures both improve upon these limitations as well as how they face their own limitations.

Covariance: Covariance is not typically used due to it's following limitations:

1. This measure depends on the units of measurement, so if the variables aren't measured the exact same way, the results would be spurious (Schober, Boer, & Schwarte, 2018; Spanos, 2019)
2. Covariance can only tell us about the relationship between 2 variables at a time. Even if there are 10 variables, variable 1 has to be compared with all 9 other variables one at a time
3. Data has to be linearly related and continuous (Delicado & Smrekar, 2009)

Correlation: This fixes the units of measurement problem of covariance, but there are many limitations to using correlation (Aggarwal & Ranganathan, 2016; Delicado & Smrekar, 2009; Genest & Neslehova, 2014; Lopez-Paz, Hennig, & Scholkopf, 2013; Schober et al., 2018; Spanos, 2019; Tjøstheim et al., 2018). In order to be able to use correlation, not even one of the following can be violated:

1. Data has to be linearly related
2. Data has to be continuous
3. No outliers can be present, not even one
4. The sample cannot be small
5. The variables must be independent

If even one of these properties are violated, erroneous attempts in using correlation will provide inaccurate and misleading results. The following measures were proposed in order to combat these limitations:

Spearman's Rank Correlation: Since correlation cannot handle ordinal data, this measure was introduced to handle such (Tjøstheim et al., 2018). Note that like correlation, this can only be used for linear data.

$$\hat{\rho}_S = \frac{n^{-1} \sum_{i=1}^{n} R_{i,X}^{(n)} R_{i,Y}^{(n)} - (n+1)^2/4}{(n^2-1)/12} \tag{13}$$

Kendall's Rank Correlation: Since correlation cannot handle ordinal data, this measure was introduced to handle such (Tjøstheim et al., 2018). Note that like correlation, this can only be used for linear data. Also note that c represents the number of concordant pairs, while d represents the number of concordant pairs.

$$\hat{\tau} = \frac{c-d}{n(n-1)/2} \tag{14}$$

Distance Correlation: This improves upon Spearman's and Kendall's rank correlations since it can handle nonlinear data and doesn't require normality (Szekely, Rizzo, & Bakirov, 2007; Tjøstheim et al., 2018).

$$dR^2(X,Y) = \frac{V^2(X,Y)}{\sqrt{V^2(X)V^2(Y)}} \tag{15}$$

Note that distance correlation assumes $V^2(X)V^2(Y) > 0$, and it is found with distance covariance, using a weighted characteristic functional.

$$V^2(X,Y;w) = \int_{\mathbb{R}^{p+q}} |\phi_{X,Y}(u,v) - \phi_X(u)\phi_Y(v)|^2 w(u,v)dudv \tag{16}$$

Gamma Coefficient: This measure was introduced to measure dependence between ordinal data (Spanos, 2019). Notice the coefficient is interpreted just like the standard correlation coefficient (i.e. if $|\gamma| = 1$ the data are perfectly associated).

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}, s.t. -1 \leq \gamma \leq 1 \tag{17}$$

This includes concordance:

$$\Pi_c = 2\sum_{i=1}^{m}\sum_{i=1}^{n}\pi_{ij}(\sum_{h>i}\sum_{k>j}\pi_{hk}) \tag{18}$$

This also includes discordance:

$$\Pi_d = 2\sum_{i=1}^{m}\sum_{i=1}^{n}\pi_{ij}(\sum_{h>i}\sum_{k<j}\pi_{hk}) \tag{19}$$

Copula: In the literature copulas are referred to as the fix for the above limitations when measuring dependence, but copulas have an important limitation too. In it's transformation, the probabilistic structure for the original random variable X is not preserved (Spanos, 2019). This means that the transformation is not monotonic, so any of the core three probabilistic assumptions can be violated which lead to erroneous inferences.

Mutual Information: Mutual information is defined as $I(X;Y) = D_{KL}(P_{(X,Y)}||P_X \otimes P_Y)$, where D_{KL} is the Kullback-Leibler divergence, $P_{(X,Y)}$ signifies the joint distribution, and P_X & P_Y signifies the marginal distributions (Kraskov, Stogbauer, & Grassberger, 2004). This measures dependence in a way that isn't constrained to linearity and real-valued random variables because it's comparing the joint distribution to the product of the marginal distributions. Intuitively mutual information measures the information that X and Y share by measuring how the knowledge of X or Y lowers uncertainty of the other random variable (Kraskov et al., 2004). Although this is certainly an improvement of correlation and other linearly constrained measures, this measure does face issues when in high-dimensional space because of the curse of dimensionality (Goldfeld & Greene, 2021).

1.3 Model Based Methods

1.3.1 Definitions

Linear Regression: The following is the most common form of a Linear Regression. The β's capture the coefficient of how much the values of each x impact y, and the ε is the error term of the estimation (Wooldridge, 2012).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \varepsilon \tag{20}$$

Here the process minimizes the sum of squared residuals $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. This aims for BLUE: a best, linear and unbiased estimator. Here best refers to relative efficiency, essentially the best estimator from the pool of linear and unbiased estimators.

Granger Causality: If x_t values provide statistically significant information on the future values of y_t, then x_t can be said to granger-cause y_t (Granger, 1969). This is supported by t-tests and F-tests. Before mentioning some limitations of granger causality, it's important to clarify here that there is a difference between correlation, correlation doesn't necessarily imply causation. There are things granger causality cannot capture including non-linear causal relationships, latent confounding effects, and instantaneous relationships. The equation below mathematically represents this by comparing conditional probabilities of the universe of available information at time t (via I_t), and that universe excluding X (via $I_t(-X)$) (Eichler, 2012):

$$\mathbb{P}[Y_{t+1}|I_t] \neq \mathbb{P}[Y_{t+1}|I_t(-X)] \tag{21}$$

AR, ARMA, ARCH & GARCH: The Autoregressive (AR) model estimates the current y_t by looking at both it's prior values.

$$AR(p) : y_t = \alpha_0 + \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \varepsilon_t \equiv \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t \tag{22}$$

The Autoregressive Moving Average (ARMA) model estimates the current y_t by looking at both it's prior values as well as the prior values of the error term (i.e. the Moving Average) (Box & Jenkins, 1976).

$$ARMA(p,q) : y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \tag{23}$$

The most common extension of the AR model is the ARCH model, it's only difference is that it accounts for conditional heteroskedasticity (Engle, 1982). Very similar to the ARCH model is the GARCH model, it's the most common extension of the ARMA model, it's simply an ARCH that includes a MA component (Bollerslev, 1986). An important extension to the GARCH is Dynamic Conditional Correlation (GARCH-DCC), which is simply a multivariate version of the GARCH model (Engle, 2002).

2 Literature

Dependence measures can be seen outside the fields of statistics and economics, it can also be seen in the fields of finance (also referred to as econophysics), neuroeconomics, as well as research in health and medical fields. These are only a few fields, this can also be seen in any field of study that runs statistical analyses on data.

2.1 Dependence and it's Issues in Various Fields

One of the most common measures of dependence regularly used in a variety of fields is correlation. In the literature, there are times the required assumptions don't hold. Remember, if even one doesn't hold, correlation is no longer an appropriate measure to use. There are so many fields, but there are two to focus on: neuroeconomics and the medical field. It is important to note that in both of these fields, there are papers that discuss the need of ensuring accuracy in using these measures (Aggarwal & Ranganathan, 2016; Botvinik-Nezer et al., 2020).

In the field of neuroeconomics (which is closely tied to experimental economics), there are some papers that erroneously fit a line for non-linear data that includes outliers. As discussed earlier, either of those will confound the basic correlation measure leading to invalid inferences (Englemann, Meyer, Ruff, & Fehr, 2019; Huettel, Stowe, Gordon, Warner, & Platt, 2016; Levy, Snell, Nelson, Rustichini, & Glimcher, 2009; Morishima, Schunk, Bruhin, Ruff, & Rehr, 2012; Weinrabe, Chung, Tymula, Tran, & Hickie, 2020). Note though that these experiments are very carefully designed and cover very interesting research questions, and researchers in the field are directly addressing this issue and other issues of methodological accuracy (Botvinik-Nezer et al., 2020). In regards to using correlation, the only thing that needs to be ensured is that correlation is used properly if it is used. This will ensure that their claims regarding dependence between variables are accurate.

Also note that there are also papers in this field that do not fall into this trap (Aimone, Ball, & King-Casas, 2016; Galvan & Peris, 2014). For example, one study uses Spearman's correlation instead of the typical correlation measure, since this better accounts for ranked data (Aimone et al., 2016). Another example shows scatterplots with outliers that have both the line and R^2 values, but it is explicitly written that these figures are for illustrative purposes only (Galvan & Peris, 2014). Explicitly writing this caveat is good because linear measures can't handle outliers (which are present) and this relationship visually doesn't even appear to be linear. Note that there is another discussion in the literature on the correlation-causation for physical processes of the brain, but this isn't discussed here since this is a separate topic of neuroscience (Ruff & Huettel, 2014).

The issue of accounting for dependence is also rampant in the research of health and medical fields that include (but are not limited to) psychology, biology, and medical researchers (i.e. anesthesiology, gastroenterology, etc.) (Aggarwal & Ranganathan, 2016; Janse et al., 2021; Schober et al., 2018). Their own journals are directly addressing this though by publishing articles that warn those in their fields about the issues of the standard correlation measure.

2.2 New Methods of Measuring Dependence

There is also a body of literature within mathematics and statistics that proposes new methods of measuring dependence to combat the limitations of prior methods.

Maximal Information Coefficient: Also known as MIC, this measure captures pairwise dependence by maximizing the mutual information (which assumes a data distribution to ignore transformations) (Nguyen, Muller, Vreeken, Efros, & Bohm, 2014; Shen, 2020; Tjøstheim et al., 2018; Yin, 2004). This measure can only capture dependence between scalar random variables, can face challenges in implementation, and has low power in finding dependence (Ding & Li, 2015; Lopez-Paz et al., 2013; Nguyen et al., 2014). In fact, distance correlation has been found to be more robust than MIC, but has it's own limitations as well (Ding & Li, 2015; Nguyen et al., 2014; Tjøstheim et al., 2018).

Maximal Correlation: This measure has a few names and variations, but they all perform multivariate maximal correlation in high dimension spaces (Nguyen et al., 2014; Shen, 2020). There are assumptions that must be properly accounted; for example, with maximum marginal correlation the choice of marginal correlation must satisfy independence properties, and there are properties of the dimension of each random variable that must also be satisfied (Shen, 2020).

Randomized Dependence Coefficient: Based on the Hirschfeld-Gebelein-Renyi Maximum Correlation Coefficient, this measure looks at the correlation of non-linear random copula projections (Lopez-Paz et al., 2013). An argued benefit is that it's easier to implement (compared to MIC), is invariant to marginal distribution transformations, and has a lower computational cost (compared to MIC). This was supported by tests on both synthetic and real-world data, and can be visualized with the following figure.

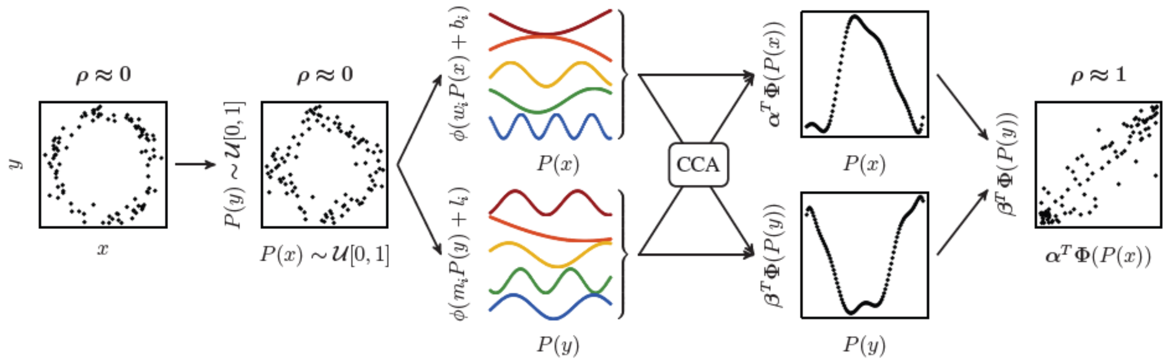


Figure 2: RDC Flowchart: This is achieved by transforming data into empirical copulas, which are then projected through randomly chosen non-linear maps, and lastly the largest correlation is found between both non-linear random projections

Predictive Power Score (PPS): This measure can practically be used in the same way as correlation would, but it's much better due to it's compatibility with non-linearly related data as well as data that is either categorical or nominal (Mai-Nguyen, Tran, Dao, & Zettsu, 2020). Also, this measures something called asymmetric correlation which means even if variable X_t is correlated with variable Y_t, variable Y_t isn't necessarily correlated with variable X_t (Levi, Karnieli, & Paz-Kagan, 2022). An example of an asymmetric relationship is between the US and one of it's states (say California); if someone lives in California you'd know they live in the US, but if someone lives in the US, you wouldn't know exactly which state they live in.

There is an important limitation to note though about this measure, this score is calculated with different evaluation metrics for different types of variables, so the score cannot be compared in a strict mathematical way (Wetschoreck, 2020). Instead, PPS should be used to find patterns, but then needs to be evaluated with a different measure (i.e. correlation, if appropriate).

Acknowledgements

Specials thanks to the PPE Center for funding this research.

References

Aggarwal, R., & Ranganathan, P. (2016). Common pitfalls in statistical analysis: The use of correlation techniques. *Perspectives in Clinical Research*, *7*(4), 187-190. Retrieved from https://doi.org/10.4103/2229-3485.192046 DOI: 10.4103/2229-3485.192046

Aimone, J., Ball, S., & King-Casas, B. (2016). It's not what you see but how you see it: Using eye-tracking to study the risky decision-making process. *Journal of Neuroscience, Psychology, and Economics*, *9*(3 & 4), 137-144. Retrieved from http://dx.doi.org/10.1037/npe0000061 DOI: 10.1037/npe0000061

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327. Retrieved from https://doi.org/10.1016/0304-4076(86)90063-1 DOI: 10.1016/0304-4076(86)90063-1

Botvinik-Nezer, R., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84-88. Retrieved from https://doi.org/10.1038/s41586-020-2314-9 DOI: 10.1038/s41586-020-2314-9

Box, G., & Jenkins, G. (1976). *Time series analysis, forecasting and control.* Holden-Day.

Buhlmann, P. (2017). High-dimensional statistics, with applications to genome-wide association studies. *EMS Surveys in Mathematical Sciences*, *4*(1), 45-75. Retrieved from https://doi.org/10.4171/emss/4-1-3 DOI: 10.4171/EMSS/x

Cardano, G. (2015). *The book on games of chance: The 16th-century treatise on probability.* Dover Publications.

Darwin, C., & Galton, F. (n.d.). *Correspondence between charles darwin and francis galton.* Retrieved from https://galton.org/letters/darwin/correspondence.htm

Delicado, P., & Smrekar, M. (2009). Measuring non-linear dependence for two random variables distributed along a curve. *Universitat Politecnica de Catalunya*. Retrieved from https://doi.org/10.48550/arXiv.2001.01095 DOI: 10.48550/ARXIV.2001.01095

Ding, A., & Li, Y. (2015). Copula correlation: An equitable dependence measure and extension of pearson's correlation. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.1312.7214 DOI: 10.48550/arXiv.1312.7214

Eichler, M. (2012). *Causal inference in time series analysis.* Wiley. Retrieved from https://doi.org/10.1002/9781119945710.ch22 DOI: 10.1002/9781119945710.ch22

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987-1007. Retrieved from https://doi.org/10.2307/1912773 DOI: 10.2307/1912773

Engle, R. (2002). Dynamic conditional correlation a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339-350. Retrieved from https://doi.org/10.1198/073500102288618487 DOI: 10.1198/073500102288618487

Englemann, J., Meyer, F., Ruff, C., & Fehr, E. (2019). The neural circuitry of affect-induced distortions of trust. *Science Advances*, *5*(3). Retrieved from https://doi.org/10.1126/sciadv.aau3413 DOI: 10.1126/sciadv.aau3413

Fancher, R. E. (2009). Scientific cousins: The relationship between charles darwin and francis galton. *American Psychologist*, *64*(2), 84–92. DOI: 10.1037/a0013339

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, *222*, 309-368. Retrieved from https://doi.org/10.1098/rsta.1922.0009 DOI: 10.1098/rsta.1922.0009

Galton, F. (1889). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, *45*, 135–145.

Galvan, A., & Peris, T. (2014). Neural correlates of risky decision making in anxious youth and healthy controls. *Depression and Anxiety*, *31*(7), 591-598. Retrieved from https://doi.org/10.1002/da.22276 DOI: 10.1002/da.22276

Genest, C., & Neslehova, J. (2014). *Statistics in action: A canadian outlook*. Chapman and Hall/CRC. Retrieved from https://doi.org/10.1201/b16597 DOI: doi.org/10.1201/b16597

Goldfeld, Z., & Greene, T. (2021). Sliced mutual information: A scalable measure of statistical dependence. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.2110.05279 DOI: 10.48550/arXiv.2110.05279

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424-438. Retrieved from https://doi.org/10.2307/1912791 DOI: 10.2307/1912791

Hald, A. (1988). On de movire's solutions of the problem of duration of play 1708–1718. *Archive for History of Exact Sciences*. DOI: 10.1007/BF00348454

Huettel, S., Stowe, J., Gordon, E., Warner, B., & Platt, M. (2016). Neural signatures of economic preferences for risk and ambiguity. *Neuron*, *49*(5), 765–775. Retrieved from https://doi.org/10.1016/j.neuron.2006.01.024 DOI: 10.1016/j.neuron.2006.01.024

Janse, R., Hoekstra, T., Jager, K., Zoccali, C., Tripepi, G., Dekker, F., & van Diepen, M. (2021). Conducting correlation analysis: important limitations and pitfalls. *Clinical Kidney Journal*, *14*(11), 2332–2337. Retrieved from https://doi.org/10.1093/ckj/sfab085 DOI: 10.1093/ckj/sfab085

Kolmogorov, A. (1933). *Foundations of probability theory*. Julius Springer.

Kraskov, A., Stogbauer, H., & Grassberger, P. (2004). Estimating mutual information. *PHYSICAL REVIEW E*, *69*(6). Retrieved from https://doi.org/10.1103/PhysRevE.69.066138 DOI: 10.1103/PhysRevE.69.066138

Laplace, P.-S. (1812). *Théorie analytique des probabilités*. Courcier Imprimeur.

Levi, N., Karnieli, A., & Paz-Kagan, T. (2022). Airborne imaging spectroscopy for assessing land-use effect on soil quality in drylands. *ISPRS Journal of Photogrammetry and Remote Sensing*, *186*, 34-54. Retrieved from https://doi.org/10.1016/j.isprsjprs.2022.01.018 DOI: 10.1016/j.isprsjprs.2022.01.018

Levy, I., Snell, J., Nelson, A., Rustichini, A., & Glimcher, P. (2009). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology*, *103*(2), 1036-1047. Retrieved from https://doi.org/10.1152/jn.00853.2009 DOI: 10.1152/jn.00853.2009

Lopez-Paz, D., Hennig, P., & Scholkopf, B. (2013). The randomized dependence coefficient. *Max Planck Institute for Intelligent Systems*. Retrieved from https://doi.org/10.48550/

arXiv.1304.7717 DOI: 10.48550/arXiv.1304.7717

Mai-Nguyen, A.-V., Tran, V.-L., Dao, M.-S., & Zettsu, K. (2020). Leverage the predictive power score of lifelog data's attributes to predict the expected athlete performance. *CEUR Workshop Proceedings*, *2696*(97), 1-13. Retrieved from https://ceur-ws.org/Vol-2696/paper_97.pdf

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C., & Rehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*(1), 73-79. Retrieved from https://doi.org/10.1016/j.neuron.2012.05.021 DOI: 10.1016/j.neuron.2012.05.021

Neyman, J., & Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Lond. A*, *231*(694-706). Retrieved from https://doi.org/10.1098/rsta.1933.0009 DOI: 10.1098/rsta.1933.0009

Nguyen, H., Muller, E., Vreeken, J., Efros, P., & Bohm, K. (2014). Multivariate maximal correlation analysis. *International Conference on Machine Learning*.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *187*, 253–318.

Pearson, K. (1914-1930). *The life, letters and labours of francis galton* (Vol. 1-3). Cambridge University Press.

Ruff, C., & Huettel, S. (2014). *Neuroeconomics*. Elsevier. Retrieved from http://dx.doi.org/10.1016/B978-0-12-416008-8.00006-1 DOI: 10.1016/B978-0-12-416008-8.00006-1

Rüschendorf, L. (2009). On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, *139*(11), 3921-3927. Retrieved from https://doi.org/10.1016/j.jspi.2009.05.030 DOI: 10.1016/j.jspi.2009.05.030

Schober, P., Boer, C., & Schwarte, L. (2018). Correlation coefficients: Appropriate use and interpretation. *ANESTHESIA ANALGESIA*, *126*(5), 1763–1768. Retrieved from https://doi.org/10.1213/ane.0000000000002864 DOI: 10.1213/ANE.0000000000002864

Schreiber, P. C. (1962). Cross-spectral analysis: An introduction. *The Journal of the Acoustical Society of America*, *34*(6), 1063–1068.

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

Shen, C. (2020). High-dimensional independence testing and maximum marginal correlation. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.2001.01095 DOI: 10.48550/ARXIV.2001.01095

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, *8*, 229-231.

Spanos, A. (2019). *Probability theory and statistical inference*. Cambridge University Press. Retrieved from https://doi.org/10.1017/9781316882825 DOI: 10.1017/9781316882825

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard UP.

Szekely, G., Rizzo, M., & Bakirov, N. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. Retrieved from https://doi.org/

10.1214/009053607000000505 DOI: 10.1214/009053607000000505

Tjøstheim, D., Otneim, H., & Støve, B. (2018). Statistical dependence: Beyond pearson's ρ. *Statistical Science*, *37*(1), 90-109. Retrieved from https://doi.org/10.1214/21-STS823 DOI: 10.1214/21-STS823

Uddin, G., Rahman, M., Hedström, A., & Ahmed, A. (2019). Cross-quantilogram-based correlation and dependence between renewable energy stock and other asset classes. *Energy Economics*, *80*, 743-759. Retrieved from https://www.sciencedirect.com/science/article/pii/S0140988319300714 DOI: https://doi.org/10.1016/j.eneco.2019.02.014

Weinrabe, A., Chung, H.-K., Tymula, A., Tran, J., & Hickie, I. (2020). Economic rationality in youth with emerging mood disorders. *Journal of Neuroscience, Psychology, and Economics*, *13*(3), 164-177. Retrieved from https://psycnet.apa.org/doi/10.1037/npe0000129 DOI: 10.1037/npe0000129

Wetschoreck, F. (2020). *Rip correlation. introducing the predictive power score.* Retrieved from https://towardsdatascience.com/rip-correlation-introducing-the-predictive-power-score-3d90808b9598

Wooldridge, J. (2012). *Introductory econometrics: a modern approach.* Cengage Learning.

Wu, W., & Mielniczuk, J. (2010). *A new look at measuring dependence.* Springer Link. Retrieved from https://link.springer.com/chapter/10.1007/978-3-642-14104-1_7 DOI: 10.1007/978-3-642-14104-1_7

Yin, X. (2004). Canonical correlation analysis based on information theory. *Journal of Multivariate Analysis*, *91*(2), 161-176. Retrieved from https://doi.org/10.1016/S0047-259X(03)00129-5 DOI: 10.1016/S0047-259X(03)00129-5