Copula Functions: Characterizing Uncertainty in Probabilistic Systems

Pranesh Kumar
Department of Mathematics and Statistics
University of Northern British Columbia
Prince George, BC V2N 4Z9, Canada
kumarp@unbc.ca

Abstract. Understanding and measuring uncertainty is central in risk analysis. Uncertainty emerges when there is less information than the total information required to describe a system and environment. Uncertainty and information are so closely associated that information provided by an experiment for instance is equal to the amount of uncertainty removed. Uncertainty prevails in several forms and various kinds of uncertainties may arise from random fluctuations, incomplete information, imprecise perception, vagueness etc. The probability theory based framework deals with the uncertainty of random phenomenon while fuzzy set concept provides an appropriate mathematical framework to deal with vagueness. We in this paper apply the concept of copula functions to characterize uncertainty associated with the probabilistic systems. Copula functions join uniform marginal distributions of random variables to form their multivariate distribution functions. Copulas are useful because they separate joint distributions into two contributions- (i) marginal distributions of each variable and (ii) copula as a measure of dependence. Several families of copulas with varying shapes and simulation programs are available providing flexibility in copula based modeling.

Keywords: Probabilistic uncertainty, Distributions functions, Copulas, Information measures, Entropy, Simulation.

1. Introduction

Uncertainty plays a vital role in our differing perceptions about the phenomena
observed around us. As our perception of the world gets more complex, the number of phenomena about which we are uncertain increases as well the uncertainty about each phenomena. To understand and decrease this uncertainty, we collect increasing amount of information. However this may cause an increase in uncertainty instead of helping to understand it. We may for instance refer to the second law of thermodynamics which states: Uncertainty in the world always tends to increase. Uncertainty is not a monolithic concept. It may appear in several forms like a probabilistic phenomenon for example occurrence or nonoccurrence of a random event or like a deterministic phenomenon where we know that the outcome is not governed by chance but we are fuzzy about the possibility of a specific outcome. In this paper we deal with the probabilistic uncertainty. Enormous growth has been witnessed with regard to the applications of information theoretic framework in physical, engineering, biological and social sciences and more so in the fields of information technology, nonlinear systems and molecular biology. Shannon [35] laid the mathematical foundation of information theory in the context of communication theory in his seminal paper: A mathematical theory of communication. He defined a probabilistic measure of uncertainty referred to as entropy. However earlier contributions in this direction have been due to Nyquist [31] and Hartley [13]. The remarkable success of Shannon's entropy measure has been primarily because it could quantify and analyze uncertainty present in the probabilistic systems. Since Shannon's work, significant contributions to the area of entropy optimizing principles and information measures have been made [18-20]. We summarize some basic results about entropy in characterizing the stochastic dependence between two discrete random variables. However, random variables could be either discrete or continuous. Generalizations to the multivariate situations are obvious.

Consider a finite real valued discrete random variable $X$ having the probability distribution $(x_i, p_i, i = 1,..m; \sum_i p_i = 1)$. The Shannon’s measure of uncertainty associated with the variable $X$ is called entropy and is defined by $H(X) = -c \sum_i p_i \log p_i$ where $c$ is an arbitrary constant. Constant $c$ is generally taken as unity and logarithm base 2 when entropy is measured in bits. The uncertainty takes the maximum value when all probabilities are equal, i.e., $p_i = 1/m$. Thus, the bounds for $H(X)$ are: $0 \leq H(X) \leq m$. Zero entropy implies that the process of generating $X$ is deterministic. Closer is $H(X)$ to 0, lesser is the uncertainty of $X$ while the value of $H(X)$ being closer to $m$ means greater uncertainty of $X$. $H(X)$ is a monotonic increasing function of $m$.

For simplicity in notations, we denote two random variables by $X$ and $Y$ with respective probability distributions $(x_i, p_i, i = 1,..m; \sum_i p_i = 1)$ and $(y_j, q_j, j = 1,..n; \sum_j q_j = 1)$ and the joint probability distribution $(x_i, y_j, p_{ij}; \sum_{ij} p_{ij} = 1)$ where
Copula functions

\( p_{ij} \neq 0 \) is the probability of a pair \((x_i, y_j)\) belonging to the rectangle \( R_i: [x_{i-1}; x_i] \times C_j: [y_{j-1}; y_j] \) following the partitioning of codomain of \( X \) and \( Y \). The joint entropy of \( X \) and \( Y \) is defined by

\[
H(X, Y) = - \sum_{ij} p_{ij} \log p_{ij}. \tag{1.1}
\]

When \( X \) and \( Y \) are independent, \( p_{ij} = p_i q_j \), \( \forall \ i = j \) and the entropy of the joint distribution equals the sum of respective entropies of \( X \) and \( Y \), i.e., \( H(X, Y) = H(X) + H(Y) \). However when they are not independent, we look for an answer to the question: How much uncertainty of \( X \) diminishes if we know that \( Y = C_j \). For more properties of entropy, we refer to Mathai and Rathie [26] and Kapur [18]. For general considerations stochastic dependence of random variables \( X \) and \( Y \) results in reducing their joint entropy. In such a case, it is relevant to introduce the conditional entropy \( H(X \mid y_j) \) which represents the amount of uncertainty of \( X \) given that \( y_j \) is observed; \( H(X \mid y_j) = - \sum_i p_{ij} \log p_{ij} \) where \( p_{ij} \) is the conditional probability of \( X \) taking a value \( x_i \) given that \( Y \) has assumed a value \( y_j \). The conditional entropy \( H(X \mid Y) \) is the amount of uncertainty of \( X \) remaining given advance knowledge of \( Y \) and is obtained by averaging \( H(X \mid y_j) \) over all \( j \) and equals to

\[
H(X \mid Y) = - \sum_{ij} p_{ij} \log p_{ij}. \tag{1.2}
\]

Similarly the conditional entropy \( H(Y \mid X) \) is defined. Conditional entropy is nonnegative and nonsymmetric. It is easily verified that \( H(X, Y) = H(Y) + H(X \mid Y) = H(X) + H(Y \mid X) \) and, therefore \( H(X, Y) \leq H(X) \) or \( H(Y) \) with equality holding if and only if \( X \) and \( Y \) are stochastically independent.

For a better understanding, if we assume \( X \) and \( Y \) are input and output respectively of a stochastic system, then \( H(X) \) represents the uncertainty of input \( X \) before output \( Y \) is observed while \( H(X \mid Y) \) is the uncertainty of input \( X \) after output \( Y \) is realized. The difference \( I(X, Y) = H(X) + H(Y) - H(X, Y) = H(X, Y) - H(X \mid Y) - H(Y \mid X) \) is called the mutual information (distance from independence) between \( X \) and \( Y \). An interesting alternative for characterizing dependence is the expression of mutual information in terms of the Kullback-Liebler divergence between joint distribution and two marginal distributions [24], defined by

\[
I(X, Y) = \sum_{ij} p_{ij} \log \left( \frac{p_{ij}}{p_i q_j} \right). \tag{1.3}
\]

where the Kullback-Liebler divergence between two probability distributions \( p \) and \( q \) is \( K(p \| q) = \sum_i p_i \log \left( \frac{p_i}{q_i} \right) \). Mutual information can also be expressed in terms of divergence between conditional and marginal distributions as
\[ I(X, Y) = \sum_j q_j \sum_i p_{i/j} \log(p_{i/j}/p_i). \] (1.4)

Mutual information thus measures the decrease in uncertainty of \( X \) caused by the knowledge of \( Y \) which is the same as the decrease in uncertainty of \( Y \) caused by the knowledge of \( X \). The measure \( I(X, Y) \) indicates the amount of information of \( X \) contained in \( Y \) or the amount of information of \( Y \) contained in \( X \). Obviously \( I(X, X) = H(X) \), the amount of information contained in \( X \) about itself.

Information Gain \( IG(X|Y) \): To transmit \( X \), how many bits on average would it save if both ends of the line knew \( Y \)? Information gain answers this question and is defined by

\[ IG(X|Y) = H(X) - H(X|Y). \] (1.5)

It is easily seen that \( I(X, Y) = I(Y, X) \) and \( I(X, Y) \geq 0 \) with equality when \( X \) and \( Y \) are stochastically independent and \( I(X, Y) \leq H(X) \). The relative information gain is defined by [22]:

\[ r(X|Y) = I(X, Y)/H(X) = 1 - H(X|Y)/H(X) = I(X, Y)/[H(X, Y) - H(Y|X)]. \] (1.6)

which shows how much information about \( Y \) diminishes the uncertainty of \( X \) relative to the initial uncertainty of \( X \). The relative information gain \( r(X|Y) \) satisfies: \( 0 \leq r(X|Y) \leq 1 \). The relative gain \( r(X|Y) \) assumes 0 value if and only if \( X \) and \( Y \) are stochastically independent. In case there is no information about which random variable influences other or which takes values first, then a symmetrical relative information gain measure is defined by

\[ R(X, Y) = 2I(X, Y)/[H(X) + H(Y)] = 2I(X, Y)/[H(X, Y) + I(X, Y)], \] (1.7)

which expresses the uncertainty from the joint distribution of \( X \) and \( Y \) to the uncertainty in case of independence. This measure \( R(X|Y) \) has the properties: \( 0 \leq R(X, Y) \leq 1 \) and \( R(X, Y) = 0 \) if and only if \( X \) and \( Y \) are stochastically independent. The measures \( r(X|Y) \) and \( R(X, Y) \) can be used to characterize the stochastic dependence between \( X \) and \( Y \). They are also useful in characterizing the dependence of qualitative variables under the hypothesis that the values of the qualitative variable cover all possibilities and their common part is empty.

In multivariate data modelling for an understanding of stochastic dependence the notion of correlation has been central. Although correlation is one of the omnipresent concepts in statistical theory, it is also one of the most misunderstood concepts. The
confusion may arise from the literary meaning of the word to cover any notion of dependence. From mathematics point of view, correlation is only one particular measure of stochastic dependence. It is the canonical measure in the world of multivariate normal distributions and in general for spherical and elliptical distributions. However empirical research in many applications indicates that the distributions of the real world seldom belong to this class. Alternatively, we collect and present ideas of copula functions which are useful in studying stochastic dependence with applications in statistical inference and simulation.

2. Copula Functions

Sklar’s theorem [35] states that any multivariate distribution can be expressed as the \( k \)-copula function \( C(u_1, \ldots, u_i, \ldots, u_k) \) evaluated at each of the marginal distributions. Copula is not unique unless the marginal distributions are continuous. Using probability integral transform, each continuous marginal \( U_i = F_i(x_i) \) has a uniform distribution on \( [0,1] \) where \( F_i(x_i) \) is the cumulative integral of \( f_i(x_i) \) for the random variable \( X_i \in (-\infty, \infty) \). The \( k \)-dimensional probability distribution function \( F \) has a unique copula representation

\[
F(x_1, x_2, \ldots, x_k) = C(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)) = C(u_1, u_2, \ldots, u_k). \tag{2.1}
\]

The joint probability density function in copula form is written as

\[
f(x_1, x_2, \ldots, x_k) = \Pi_{i=1}^k f_i(x_i) \times c(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)), \tag{2.2}
\]

where \( f_i(x_i) \) is each marginal density and coupling is provided by copula density

\[
c(u_1, u_2, \ldots, u_k) = \frac{\partial^k C(u_1, u_2, \ldots, u_k)}{\partial u_1 \partial u_2 \ldots \partial u_k}, \tag{2.3}
\]

if it exists. In case of independent random variables, copula density \( c(u_1, u_2, \ldots, u_k) \) is identically equal to one. The importance of the above equation \( f(x_1, x_2, \ldots, x_k) \) is that the independent portion expressed as the product of the marginals can be separated from the function \( c(u_1, u_2, \ldots, u_k) \) describing the dependence structure or shape. The dependence structure summarized by a copula is invariant under increasing and continuous transformations of the marginals. This means that suppose we have a probability model for dependent insurance losses of various kinds. If our interest now changes to model the logarithm of these losses, the copula will not change, only the marginal distributions will change.

The simplest copula is independent copula
\[ \Pi: = C(u_1, u_2, \ldots, u_k) = u_1 u_2 \ldots u_k, \quad (2.4) \]

with uniform density functions for independent random variables. Another copula example is the Farlie-Gumbel-Morgenstern (FGM) \[12\] bivariate copula. The general system of FGM bivariate distributions is given by
\[
F(x_1, x_2) = F_1(x_1) \times F_2(x_2)[1 + \rho(1 - F_1(x_1))(1 - F_2(x_2))], \quad (2.5)
\]
and copula associated with this distribution is a FGM bivariate copula
\[
C(u, v) = uv[1 + \rho(1 - u)(1 - v)].
\]

An empirical copula may be estimated by considering \(N\) pairs of data \(\{(x_{1,t}, x_{2,t})\}_{0 < t \leq N}\) by
\[
C(n/N, m/N) = \sum t 1_{[r_{t,1} \leq n, r_{t,2} \leq m]}, \quad (2.6)
\]
where \(r_{t,1}\) and \(r_{t,2}\) are the rank statistics of \(\{x_{1,t}\}_t\) and \(\{x_{2,t}\}_t\) respectively. However empirical copula has the drawback that it requires significant number of sample data for an accurate estimation. Hence parametric copulas are preferred.

Many copulas exist and a very popular class of copulas is Archimedean copulas which has a simple form and models a variety of dependence structures. Most of the Archimedean copulas have closed-form solutions. To define an Archimedean copula, let be a continuous strictly decreasing convex function from \([0,1]\) to \([0, \infty]\) such that \(c(1) = 0\) and \(c(0) = \infty\). Let \(\tau^{-1}\) be the pseudo inverse of \(c\). Then a \(k\)-dimensional Archimedean copula is
\[
C(u_1, u_2, \ldots, u_k) = \tau^{-1}[u_1 + \ldots + u_k], \quad (2.7)
\]
The function \(c\) is known as a generator function. Any generator function satisfying \(c(1) = 0\); \(\lim_{x \to 0} c(x) = \infty\); \(c'(x) < 0\); \(c''(x) > 0\) will result in an Archimedean copula. For example \[2\], generator function \(c(t) = (t^{-\theta} - 1)/\theta, \theta \in (-1, \infty) \setminus \{0\}\) results in the bivariate Clayton copula \(C(u_1, u_2) = \max[(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, 0]\).

The copula parameter \(\theta\) controls the amount of dependence between \(X_1\) and \(X_2\).

The Fréchet-Hoeffding bounds for copulas \[7\]: The lower bound for \(k\)-variate copula is
\[
W(u_1, u_2, \ldots, u_k): = \max\{1 - n + \sum_i u_i, 0\} \leq C(u_1, u_2, \ldots, u_k). \quad (2.8)
\]
And the upper bound for \(k\)-variate copula
Copula functions

\[ C(u_1, u_2, \ldots, u_k) \leq \min_{i \in \{1, 2, \ldots, k\}} u_i =: M(u_1, u_2, \ldots, u_k). \quad (2.9) \]

For all copulas, inequality \( W(u_1, \ldots, u_k) \leq C(u_1, \ldots, u_k) \leq M(u_1, \ldots, u_k) \) is satisfied. This inequality is well known as the Fréchet-Hoeffding bounds for copulas. Further, \( W \) and \( M \) are copulas themselves. It may be noted that the Fréchet-Hoeffding lower bound is not a copula in dimension \( k \geq 2 \). Copulas \( M, W \) and \( \Pi \) have important statistical interpretations \[28\]. Given a pair of continuous random variables \((X_1, X_2)\), (i) copula of \((X_1, X_2)\) is \( M(u_1, u_2) \) if and only if each of \( X_1 \) and \( X_2 \) is almost surely increasing function of the other; (ii) copula of \((X_1, X_2)\) is \( W(u_1, u_2) \) if and only if each of \( X_1 \) and \( X_2 \) is almost surely decreasing function of the other and (iii) copula of \((X_1, X_2)\) is \( \Pi(u_1, u_2) = u_1 u_2 \) if and only if \( X_1 \) and \( X_2 \) are independent.

Three commonly used rank based concordance measures are Kendall's \( \tau \), Spearman's \( \rho \) and Gini's index \( \gamma \) which could be expressed in terms of copulas \[33\]:

\[ \tau = 4 \int_{I^2} C(u_1, u_2) dC(u_1, u_2) - 1, \quad (2.10) \]

\[ \rho = 12 \int_{I^2} u_1 u_2 dC(u_1, u_2) - 3, \quad (2.11) \]

\[ \gamma = 2 \int_{I^2} (|u_1 + u_2 - 1| - |u_1 - u_2|) dC(u_1, u_2). \quad (2.12) \]

It may however be noted that the Pearson's linear correlation coefficient can not be expressed in terms of copula.

The tail dependence index of a multivariate distribution describes the amount of dependence in the upper right tail or lower left tail of the distribution and can be used to analyze the dependence among extreme random events. Tail dependence describes the limiting proportion that one margin exceeds a certain threshold given that the other margin has already exceeded that threshold. Joe \[16\] defines the tail dependence: If a bivariate copula \( C(u_1, u_2) \) is such that

\[ \lambda_U:= \lim_{u \to 1} [(1 - 2u + C(u, u))/(1 - u)], \quad (2.13) \]

exists, then \( C(u_1, u_2) \) has upper tail dependence for \( \lambda_U \in (0, 1] \) and no upper tail dependence for \( \lambda_U = 0 \). Similarly lower tail dependence in terms of copula is defined

\[ \lambda_L:= \lim_{u \to 0} [C(u, u)/u]. \quad (2.14) \]
Copula has lower tail dependence for \( \lambda_L \in (0,1) \) and no lower tail dependence for \( \lambda_L = 0 \). This measure is extensively used in extreme value theory. This is the probability that one variable is extreme given that other is extreme. Tail measures are copula-based and copula is related to the full distribution via quantile transformations, i.e.,

\[
C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad (2.15)
\]

for all \( u_1, u_2 \in (0,1) \) in the bivariate case.

Copula Simulation: Simulation in statistical data analysis has a pivotal role in replicating and analyzing data. Copulas can be applied in simulation and Monte Carlo studies. Johnson [17] discusses methods to generate a sample from a given joint distribution. One such method is a recursive simulation using the univariate conditional distributions. The conditional distribution of \( X_j \) given first \( i - 1 \) components is

\[
c(u_i|u_1, \ldots, u_{i-1}) = \frac{\partial^{i-1}c(u_1, \ldots, u_i)}{\partial u_1 \cdots \partial u_{i-1}} / \frac{\partial^{i-1}c(u_1, \ldots, u_{i-1})}{\partial u_1 \cdots \partial u_{i-1}}. \quad (2.16)
\]

For \( k \geq 2 \), procedure is as follows: (i) Select a random number \( u_i \) from Uniform (0,1) distribution and (ii) Simulate a value \( u_k \) from \( c(u_k|u_1, \ldots, u_{k-1}) \), \( k = 2, 3, \ldots \). Some important contributions in the area of copulas are listed in the references.

3. Information-Theoretic Measures Using Copula Functions

The joint entropy \( H(X,Y) \) associated with the joint distribution of \( X \) and \( Y \) using copula density function \( c(u_1, u_2) \) from (2.3) can be expressed

\[
H(X,Y) = - \sum_{i,j} c(u_1, u_2) \log c(u_1, u_2). \quad (3.1)
\]

The conditional entropy \( H(X|Y) \) expressed in terms of conditional copula density function \( c(u_1|u_2) \) from (2.16) is

\[
H(X|Y) = - \sum_{i,j} c(u_1, u_2) \log c(u_1|u_2). \quad (3.2)
\]

The mutual information (distance from independence) \( I(X,Y) \) between \( X \) and \( Y \) using copula functions is expressed by

\[
I(X,Y) = - \sum_{i,j} c(u_1, u_2) \log[c(u_1, u_2)/\{c(u_1|u_2) \times c(u_2|u_1)\}]. \quad (3.3)
\]

The relative information gain \( r(X|Y) \) in terms of copula functions
Copula functions

\[ r(X|Y) = \frac{\sum_{ij} c(u_1, u_2) \log[c(u_1, u_2)/c(u_1|u_2) \times c(u_2|u_1)]}{\sum_{ij} c(u_1, u_2) \log[c(u_1, u_2)/c(u_2|u_1)]}, \quad (3.4) \]

and the symmetrical relative information gain

\[ R(X, Y) = \frac{2 \sum_{ij} c(u_1, u_2) \log[c(u_1, u_2)/c(u_1|u_2) \times c(u_2|u_1)]}{\sum_{ij} c(u_1, u_2) \log[c^2(u_1, u_2)/c(u_1|u_2) \times c(u_2|u_1)]}. \quad (3.5) \]

The calculations involved in evaluating these expressions may be cumbersome depending upon the copula functions and the marginal probability distributions. Kovacs [22] has proposed an alternative way by expressing probabilities of a pair \((x_i, y_j)\) belonging to the rectangle \(R_i: [x_{i-1}^*, x_i^*] \times C_j: [y_{j-1}^*, y_j^*]\) in terms of associated copula function \(C(u_1, u_2) = C(u, v)\) as

\[ p_{ij} = \int_{x_{i-1}}^{x_i} \int_{y_{j-1}}^{y_j} f(x, y) dx dy = \int_{u_{i-1}}^{u_i} \int_{v_{j-1}}^{v_j} c(u, v) du dv, \quad (3.6) \]

\[ p_i \frac{j}{j} = \frac{\int_{u_{i-1}}^{u_i} \int_{v_{j-1}}^{v_j} c(u, v) du dv}{\int_{v_{j-1}}^{v_j} v_{j-1}}, \quad (3.7) \]

where \(u_i^* = F_i(x_i^*)\) and \(v_j^* = F_j(y_j^*)\). The integrals appearing in (3.6) and (3.7) are expressed in terms of copula by

\[ \int_{u_{i-1}}^{u_i} \int_{v_{j-1}}^{v_j} \frac{\partial^2 C(u, v)}{\partial u \partial v} du dv = C(u_i^*, v_j^*) - C(u_{i-1}^*, v_j^*) - C(u_i^*, v_{j-1}^*) + C(u_{i-1}^*, v_{j-1}^*). \quad (3.8) \]

It is easy to evaluate information measures in (3.1) - (3.5) using (3.8) because \(C(u, v)\) need to be evaluated at the points of partition only.

4. Mutual Information in terms of Marshall-Olkin Copula

The one parameter Marshall-Olkin copula is defined by

\[ C(u, v) = \min(u^{1-\theta} v, uv^{1-\theta}), \quad u, v, \theta \in (0,1]. \quad (4.1) \]

The copula density function \(c(u, v)\) is
The copula parameter $\theta$ in terms of Kendall's $\tau$ has a simple expression

$$\theta = \frac{2\tau}{1+\tau} \quad (4.3)$$

Mercier [27] has shown that the mutual information $I(X,Y)$ is the entropy of the copula $C(u,v)$ itself whatever the marginal distributions may be. Using one parameter Marshall-Olkin copula $C(u,v)$, they have derived the explicit algebraic expression of $I(X,Y)$ as

$$I(X,Y) = -2^{\frac{1-\theta}{2-\theta}} \left[ \log(1 - \theta) + \frac{\theta}{2-\theta} \right], \quad (4.4)$$

or in terms of Kendall's $\tau$

$$I(X,Y) = -(1 - \tau) \left[ \tau + \log \left( \frac{1-\tau}{1+\tau} \right) \right]. \quad (4.5)$$

Figure 1 shows the behaviour of the joint entropy $I(X,Y)$ for the range of values of the dependence parameter $\theta \in (0,1]$. This parametrization of joint entropy based on one copula parameter is much more accurate than the correlation parameter while keeping the same level of computational complexity.

5. Examples

We consider two examples to illustrate applications of the information-theoretic uncertainty measures- one univariate and other bivariate distributions.

Example 1. Benford's Law is a powerful and relatively simple tool for pointing suspicion at frauds, embezzlers, tax evaders, sloppy accountants and even computer bugs. The income tax agencies of several nations and several states are using detection software based on Benford's Law, as are a score of large companies and accounting businesses.
Dr. Frank Benford, a physicist at the General Electric Company, noticed that pages of logarithms corresponding to numbers starting with the numeral 1 were much dirtier and more worn out than other pages. He thought that it was unlikely that physicists and engineers had some special preference for logarithms starting with 1. He therefore embarked on a mathematical analysis of 20,229 sets of numbers, including such wildly disparate categories as the areas of rivers, baseball statistics, numbers in magazine articles and the street addresses of the first 342 people listed in the book "American Men of Science." All these seemingly unrelated sets of numbers followed the same first-digit probability pattern as the worn out pages of logarithm tables suggested. In all cases, the number 1 turned up as the first digit about 30 percent of the time, more often than any other. He derived a formula to explain this phenomenon. If absolute certainty is considered as 1 and absolute impossibility as 0, then the leading digit $d \in [1, b - 1]$ in base $b$ occurs with probability $P(d) = \log_b [1 + \frac{1}{d}]$. This quantity is exactly the space between $d$ and $d + 1$ in a logarithmic scale. In base 10, the leading digits 1 through 9 have the following distribution:

<table>
<thead>
<tr>
<th>$d$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>0.301</td>
<td>0.176</td>
<td>0.125</td>
<td>0.097</td>
<td>0.079</td>
<td>0.067</td>
<td>0.058</td>
<td>0.051</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Figure 1. Mutual Information and Dependence Parameter.
The entropy as a measure of equality of digits 1 to 9 is
\[
H(d) = -\sum_i p_i \log_{10} p_i = 0.87 \text{ dits/digit}
\]
and maximum entropy
\[
H_{\text{max}}(d) = \log_{10} 9 = 0.95 \text{ dits/digit}.
\]
Thus, the uncertainty in the distribution of digits in the table is less than the maximum possible uncertainty. This reduction in uncertainty is due to the information available that all digits in the table do not occur in the same proportion.

Example 2. The viscosity \((Y\text{ in centistokes } @ 100^\circ\text{C})\) of a polymer is related to the reaction temperature \((X_1 \text{ in } ^\circ\text{C})\) and catalyst feed rate \((X_2 \text{ in lb/h})\). An experiment was conducted to model viscosity, reaction temperature and catalyst feed rate and the following results were obtained [28; page 393]:

<table>
<thead>
<tr>
<th>Viscosity</th>
<th>Temperature</th>
<th>Catalyst Feed Rate</th>
<th>Viscosity</th>
<th>Temperature</th>
<th>Catalyst Feed Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2256</td>
<td>80</td>
<td>8</td>
<td>2364</td>
<td>94</td>
<td>12</td>
</tr>
<tr>
<td>2340</td>
<td>93</td>
<td>9</td>
<td>2379</td>
<td>93</td>
<td>11</td>
</tr>
<tr>
<td>2426</td>
<td>100</td>
<td>10</td>
<td>2440</td>
<td>97</td>
<td>13</td>
</tr>
<tr>
<td>2293</td>
<td>82</td>
<td>12</td>
<td>2364</td>
<td>95</td>
<td>11</td>
</tr>
<tr>
<td>2330</td>
<td>90</td>
<td>11</td>
<td>2404</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>2368</td>
<td>99</td>
<td>8</td>
<td>2317</td>
<td>85</td>
<td>12</td>
</tr>
<tr>
<td>2250</td>
<td>81</td>
<td>8</td>
<td>2309</td>
<td>86</td>
<td>9</td>
</tr>
<tr>
<td>2409</td>
<td>96</td>
<td>10</td>
<td>2328</td>
<td>87</td>
<td>12</td>
</tr>
</tbody>
</table>

The Kendall's rank correlation \(\tau\) between \(Y\) and \(X_1\) is 0.8186 and between \(Y\) and \(X_2\) is 0.0903. Marshall-Olkin copula parameter \(\theta\) associated with \(Y\) and \(X_1\) is estimated as 0.9 and with \(Y\) and \(X_2\) as 0.1657. Thus, viscosity in this experiment is highly related with the reaction temperature however weekly related with the catalyst feed rate. In base \(e\) (natural), the mutual information \(I(X_1, Y) = 0.2697\) and \(I(X_2, Y) = 0.0826\). The decrease in uncertainty of \(Y\) caused by the knowledge of \(X_1\) is more than the decrease caused by the knowledge of \(X_2\). Thus, the amount of information of \(Y\) contained in \(X_1\) is higher than the information of \(Y\) contained in \(X_2\). The reaction temperature is a significant variable in modeling and predicting the viscosity variable.

Acknowledgments. This work was supported by the author's discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Copula functions


Received: October, 2010