Mortality Rate Modeling of Joint Lives and Survivor Insurance Contracts Tested by A Novel Unilateral Dependence Measure

Yvonne Chueh
Department of Mathematics
Central Washington University
Ellensburg, USA
Email: chueh@cwu.edu

Angel Caţaron
Electronics and Computers Department
Transylvania University of Braşov
Romania
Email: cataron@unitbv.ro

Răzvan Andonie
Computer Science Department
Central Washington University
Ellensburg, USA
and
Electronics and Computers Department
Transylvania University of Braşov
Romania
Email: andonie@cwu.edu

***Abstract*—Recently, nonsymmetric measures of dependence have started to attract attention, and several continuous entropy-like nonsymmetric dependence measures have been proposed. Based on Onicescu's information energy, we have introduced in previous work a nonsymmetric dependence measure between two discrete random variables. In the present paper, we analyze the continuous version of this measure. We deduct that there are important differences when switching from the discrete to the continuous measure. Then we apply this continuous unilateral dependence measure to a real-world challenge: mortality rate modeling for life insurance industry. We consider joint male-female pairs (married, but not necessarily from the same family) belonging to the same policy group, and analyze the unilateral interactions between the male-female mortality data samples for the purpose of stochastic model testing and validation of risk variables in the insurance world.**

I. Introduction

Measuring the statistical dependence nonparametric relationships between random variables has been a research topic in statistics and information theory for a long time, with applications in many fields, including data mining and machine learning. This paper intends to broaden its application to more data driven fields including actuarial science. There are two strategies one can adopt when studying the relationship between two random variables: the first is to measure their interdependence thought as a mutual attribute and the second is to measure how much one system depends on the other. In the first case we have symmetric (bilateral) measures of dependence, whereas in the second case we have nonsymmetric (unilateral) measures.

A. Symmetric measures

Rényi [15] gave a set of seven postulates which a measure of symmetric nonparametric dependence for two continuous random variables should satisfy on a given probability space. Rényi considered six dependence measures which satisfy some or all seven postulates. Later, Bell [5] compared several dependence measures and proposed postulate modifications. Other symmetric dependence measures were proposed in [16].

The most common symmetric dependence measure in information theory is the mutual information (MI) between two random variables X and Y: $MI(X,Y) = H(X) + H(Y) - H(X,Y) = MI(Y,X)$, which measures the dependence using Shannon's entropy. Since it is a symmetric function, it measures simultaneously the dependence of one random variable by the other and vice versa.

Sometimes, it is computationally attractive to use one of the generalized forms of the MI. Such a form is the Rényi divergence measure, which uses Rényi‘s quadratic entropy instead of the Shannon entropy. The MI and Rényi's divergence measure are equivalent but only in the limit $\alpha = 1$, where α is the order of Rényi‘s divergence measure [14].

B. Nonsymmetric measures and our previous work

Recently, nonsymmetric measures of dependence have started to attract attention, and several continuous entropy-like nonsymmetric dependence measures have been proposed [11]. Similar to Rényi's postulates for symmetric measures, Li has recently introduced six postulates a nonsymmetric continuous dependence measure should satisfy [11].

Information measures can also refer to certainty (not only to uncertainty, like Shannon's entropy), and probability can be considered as a measure of certainty. More general, any monotonically growing and continuous probability function can be considered as a measure of certainty. Onicescu's information energy (IE) [13] is a special case of Van der Lubbe *et al.* certainty measure [18] and was interpreted by several authors as a measure of expected commonness, a measure of average certainty, or as a measure of concentration, and is not related to physical energy.

Based on the IE, we have introduced in previous work the following nonsymmetric dependence measure [4]: $o(X,Y) =$

$IE(X|Y) - IE(X)$, where $IE(X) = \sum_{k=1}^{n} p_k^2$ is the IE of a discrete random variable X with probabilities p_k, and $IE(X|Y)$ is the conditional informational energy between variables X and Y. We have applied this measure to pattern classification and feature weighting problems [2], [1], [3], [6] as an alternative to the MI.

Later, we considered the continuous version of the $o(X,Y)$ measure, in the context of probability density estimation. *Density estimation* is the construction of an estimate of the density function from the observed data [17]. We refer here only to nonparametric estimation. Although it is assumed that the distribution has the probability density f, the data is allowed to speak for themselves in determining the estimate of f more than would be the case if f were constrained to fall in a given parametric family. Our goal is to compute the $o(X,Y)$ measure from the available dataset, for random variables X and Y with unknown densities. This corresponds to a regression (function approximation) learning problem. We showed how to estimate the $o(X,Y)$ measure from an available sample set of discrete or continuous variables using the kNN estimation and applied this to several real-world problems [7], [8].

C. Our contribution

Using both the discrete and the continuous versions of the $o(X,Y)$ measure made us reflect on the differences between their mathematical properties. We noticed that there are important differences when switching from discrete to continuous. It is usually more convenient to use continuous mathematics but, in practice, we often have to go back to the discrete case because we work with discrete data samples drawn from unknown distributions. Nevertheless, when our discrete data fits well with a known probability density function, the continuous approach can be considered. This gave us the motivation for the present work.

We derive properties of the continuous $o(X,Y)$ measure and relate them to Li's postulates. We discuss case studies and examples showing how this measure can be applied. We compare the properties of the discrete and continuous versions of $o(X,Y)$.

Finally, we apply the continuous $o(X,Y)$ measure to risk modeling for joint lives and survivor insurance and annuity contracts using a real-world dataset extracted from the database of the Society of Actuaries[1], with the contract data from major insurance companies in the USA. We aim to use our results for screening and selecting essential mortality model predictors.

The structure of the paper is the following. Section II is an overview of the properties of the discrete IE and the discrete $o(X,Y)$ measure, with results taken from previous papers. In addition, we discuss here how the Li's postulates are satisfied. Section III introduces the properties of the continuous IE. In Section IV we derive the properties of the continuous $o(X,Y)$ measure and relate them to Li's postulates. In Section V we apply the continuous dependence on life expectancy and mortality data which is sometimes used to assess the risk in insurance industry. Section VI concludes with some final remarks.

[1] www.soa.org

II. Properties of the discrete $o(X,Y)$ measure

In this section we review the properties of the discrete IE and the $o(X,Y)$ measure.

The discrete IE has the following properties (see [13]):

a) The value range for the discrete $IE(X)$ is between $1/n$ and 1. The value is $1/n$ iff all probabilities are equal and 1 iff one probability is 1.

b) The IE is invariant with respect to any bijective function g applied to X: $IE(g(X)) = IE(X)$. Therefore, the discrete IE is also invariant to shifting and scaling.

c) $IE(X)$ decreases when $H(X)$ increases.

The IE decreases in direct proportion with the raising of uniformity, disorder, or indetermination.

The $o(X,Y)$ measures the "additional" average certainty (or information) of X occurring under the condition that Y has already or simultaneously occurred (or is certain) "over" the average certainty of X when the certainty (or information) of Y is not available. Thus, $o(X,Y)$ can be regarded as an indicator of the unilateral dependence of X upon Y.

Based on the previous properties, we proved in [4] the following properties for the discrete $o(X,Y)$ measure:

i) o is not symmetrical with respect to its arguments.

ii) $o(X,Y) \geq 0$ and equality holds iff X and Y are independent. This results from the property: $IE(X|Y) \geq IE(X)$ with equality iff X and Y are independent.

iii) $o(X,Y) \leq 1 - IE(X)$ and equality holds iff X is completely dependent on Y.

Since $IE(X)$ is invariant with respect to any bijective function g applied to X, we also have: $o(g(X),Y) = o(X,Y)$.

According to Li, a nonsymmetric continuous dependence measure $R(X,Y)$ should satisfy [11] the following postulates:

P1: $R(X,Y)$ is defined for all non-constant continuous random variables X, Y;

P2: $R(X,Y)$ may be not equal to $R(Y,X)$;

P3: $0 \leq R(X,Y) \leq 1$;

P4: $R(X,Y) = 0$ iff X, Y are independent;

P5: $R(X,Y) = 1$ iff $Y = f(X)$ almost surely for a Borel-measurable function f;

P6: If g is a Borel-measurable bijection on $\Re$, then $R(g(X),Y) = R(X,Y)$.

The $o(X,Y)$ measure, defined for all discrete random variables X and Y, satisfies postulates **P1**, **P2**, **P4**, and **P6**. **P5** is not satisfied.

Postulate **P3** is satisfied only in a more relaxed form: The $o(X,Y)$ dependence measure has the value range $0 \leq o(X,Y) \leq 1 - IE(X)$, from independence of X and Y, to complete dependence of X on Y.

III. Properties of the continuous IE

We introduce now the main properties of the continuous IE, omitting the proofs. We will use these results in the next sections.

For a continuous random variable X with probability density function $f(x)$, the IE is [13], [10]:

$$IE(X) = \int_{-\infty}^{+\infty} f^2(x)dx \quad (1)$$

With respect of shifting and scaling, we have the following results.

Theorem 1. $IE(X + c) = IE(X)$.

This means that the IE is invariant to translations.

Theorem 2. $IE(cX) = \frac{1}{|c|} IE(X)$.

Scaling X changes $IE(X)$: The energy is reduced if $|a| > 1$; otherwise, the energy is increased. The effect of scaling is illustrated in Fig. 2.

For two continuous random variables X and Y, with their joint probability density function $f(x, y)$ we have:

$$IE(X|Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)f(x|y)dy\, dx \quad (2)$$

Similar to the discrete case, we have the following property (given here without our proof):

Theorem 3. *$IE(X|Y) \geq IE(X)$ with equality iff X and Y are independent.*

As a special important case, the IE of a multi-variate normal continuous distribution can be computed according to the following

Theorem 4. $IE(X_1, X_2, \ldots, X_n) = IE(\mathcal{N}_n(\mu, K)) = \frac{1}{\sqrt{2}(\sqrt{2\pi})^n |K|^{\frac{1}{2}}}$.

The range of values for the continuous $IE(X)$ is different than for the discrete case: We have $IE(X) > 0$, but it is possible, as we will see in an example, that $IE(X) > 1$. Also, using normal distribution examples, it can be shown that there are no theoretical uniform bound for the $IE(X)$ in continuous density models. This can be extended to joint and conditional probability density functions, resulting that $IE(X|Y)$ is also not bounded by a constant. All we can say is that the upper bounds happen at the mode where the PDFs have their peak values.

Let us highlight the differences between the IE and Shannon's entropy when applied to a uniform distribution on the interval $(0, a)$:

$$IE(X) = \int_0^a \left(\frac{1}{a}\right)^2 dx = \left(\frac{1}{a}\right)^2 a = \frac{1}{a},$$

where a is the volume of the support and is always nonnegative. We know [9] that the entropy of the continuous uniform distribution defined on the interval $(0, a)$ is $H(X) = \log a$. Hence, we remark that the continuous entropy moves into the opposite direction to the corresponding IE as the distribution parameter changes its value: While $H(X)$ increases when a or σ increase, $IE(X)$ decreases when a or σ increase.

Example of IE for a scaled random variable

Given the following Pareto probability density function function of X:

$$f(x) = \frac{\alpha\theta^\alpha}{(x + \theta)^{\alpha+1}} = \frac{3 \cdot 11^3}{(x + 11)^4},$$

where X follows a Pareto distribution with parameters $\alpha = 3$ and $\theta = 11$.

Given a scaled random variable Y and $Y = 2X$, the probability density function of Y is given as follows:

$$g(y) = \frac{\alpha\theta^\alpha}{(x + \theta)^{\alpha+1}} = \frac{3 \cdot 22^3}{(x + 22)^4}.$$

By the property of Pareto distribution, when a Pareto variable is scaled by a constant, the resulting distribution's scale parameter will be scaled by the same constant accordingly, since θ is a scale parameter. Thus, $Y = 2X$ must follow the Pareto distribution with the parameters $\alpha = 3$ and $2\theta = 22$. We obtain:

$$IE(X) = \int f^2(x)dx = \frac{9}{77} \approx 0.11688$$
$$IE(Y) = IE(2X) = \int g^2(y)dy = \frac{9}{154} \approx 0.05442.$$

Using Monte Carlo simulation, we generated 120 data points from a Pareto population with the same parameter values, and them based on the same sample data we fit the data to a Pareto density curve using the maximum likelihood estimation (MLE) for the parameter estimates: $\hat{\alpha} = 3.950149718$ and $\hat{\theta} = 11.89175072$. We obtain $IE(X) = 0.147439115$ and $\frac{1}{2}IE(X) = 0.073719557$. We then scale the sample data by a constant $c = 2$. We fit the scaled data to another Pareto density curve using the MLE parameter estimates $\hat{\alpha} = 3.950303111$ and $\hat{\theta} = 23.78465538$ to compute the density function values and the IE, which is $IE(Y) = IE(2X) = 0.073719354$. As expected, we obtained a ratio of approximately $\frac{1}{2}$ between IE(X) and IE(Y).

The above example validates the relation of a scale random variable and its resulting discrete and continuous random variables, as realized by our numerical example.

IV. Properties of the Unilateral Dependence Measure

For two continuous random variables X and Y with their joint probability density function $f(x, y)$ we have:

$$o(X, Y) = IE(X|Y) - IE(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)f(x|y)dy\, dx - \int_{-\infty}^{+\infty} f^2(x)dx \quad (3)$$

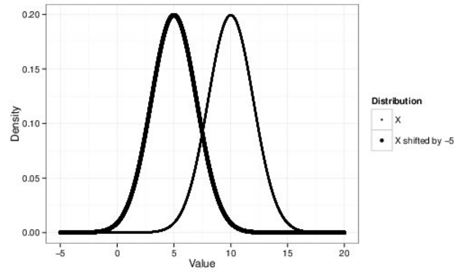

Fig. 1. The effect of shifting a normal distribution. The IE of the distribution $\mathcal{N}(10,2)$ is approximately 0.141, and the IE of its shifted version $\mathcal{N}(5,2)$ has the same value.

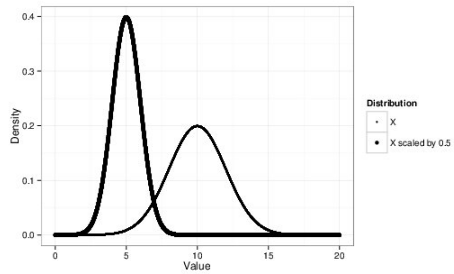

Fig. 2. The effect of scaling a normal distribution. The IE of the distribution $\mathcal{N}(10,2)$ is approximately 0.141, and the IE of its pair $\mathcal{N}(5,1)$ scaled by factor 0.5 is approximately 0.282.

We will introduce the properties of this measure, omitting the proofs. We discuss how the continuous $o(X,Y)$ verifies Li's postulates.

The measure $o(X,Y)$ quantifies the unilateral dependence characterizing X with respect to Y and corresponds to the amount of information contained in Y about X. The information is measured by using probability density. Intuitively, $o(X,Y)$ measures the expected conditional probability density deducting the expected marginal probability density. The difference, $o(X,Y)$, gives the expected additional conditional density due to the information of Y.

Theorem 5. *$o(X,Y) \neq o(Y,X)$ unless $f(x)=f(y)$ almost everywhere, which means $X=Y$.*

Theorem 6. *$o(X,Y) \geq 0$ with equality iff X and Y are independent.*

With respect to scaling and shifting, we have the following properties:

Theorem 7. $o(X,Y)=o(X+c,Y)$.

Theorem 8. $o(cX,Y)=\frac{1}{|c|}o(X,Y)$.

In particular, we have:

Theorem 9. $o(X+c,X)=o(X,X+c)=o(cX,X)=o(X,cX)=1$.

This means that X, cX, and $X+c$ are completely dependent, which is exactly what we expect.

Consequently, the continuous $o(X,Y)$ measure satisfies the following postulates: **P1**, **P2** (this results from Theorem 5), and **P4** (this results from Theorem 6). **P5** is not satisfied. **P6** is also not satisfied (for instance, for scaling - see Theorem 8).

Postulate **P3** is partially satisfied: We only have: $o(X,Y) \geq 0$ with equality iff X and Y are independent.

As an example, let us compute the unilateral dependence measure between X and Y, two Gaussian random variables with correlation σ.

We may conclude that the bivariate (X,Y) follows a bivariate Gaussian distribution and it implies by shifting the mean to zero, $(X,Y) \sim \mathcal{N}(0,K)$, (X,Y) follows a bivariate normal distribution with the mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and the covariance matrix $K=\begin{pmatrix} \sigma_X^2 & \sigma\sigma_{XY} \\ \sigma\sigma_{XY} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \sigma\sigma_X\sigma_Y \\ \sigma\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$.

The conditional random variable $X|Y \sim \mathcal{N}(\mu_{X|Y}, \sigma_{X|Y})$.

$$\mu_{X|Y} = \mu_X + \sigma\frac{\sigma_X}{\sigma_Y}(Y-\mu_Y) = 0 + \sigma\frac{\sigma_X}{\sigma_Y}(Y-0)$$

$$\sigma_{X|Y}^2 = \sigma_X^2(1-\sigma^2)$$

$$IE(X|Y) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_{X|Y}} = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_X\sqrt{1-\sigma^2}}$$

$$IE(X) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_X}$$

$$IE(Y|X) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_{Y|X}} = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_Y\sqrt{1-\sigma^2}}$$

$$IE(Y) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_Y}$$

$$o(X,Y) = IE(X|Y) - IE(X) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_X}\left(\frac{1}{\sqrt{1-\sigma^2}} - 1\right)$$

$$o(Y,X) = IE(Y|X) - IE(Y) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sigma_Y}\left(\frac{1}{\sqrt{1-\sigma^2}} - 1\right)$$

If $\sigma = 0$, then $o(X,Y) = o(Y,X) = 0$. If $\sigma = \pm 1$, then $o(X,Y) = o(Y,X) = +\infty$, when $-1 < \sigma < 1$.

If $\sigma_X > \sigma_Y$, then $o(X,Y) < o(Y,X)$.

If $\sigma_X < \sigma_Y$, then $o(X,Y) > o(Y,X)$.

If $\sigma_X = \sigma_Y$, then $o(X,Y) = o(Y,X)$.

In a multi-variate Gaussian, $\sigma = 0$ implies independence: the unilateral dependence measure is 0. $\sigma = \pm 1$ indicates that X and Y are perfectly linearly correlated: the unilateral dependence measure is infinite.

The standard deviation of the marginal X, σ_X relative to σ_Y, determines the unilateral dependence measure.

$\frac{\sigma_X}{\sigma_Y} = 1$, X and Y are equally dependent on the other.
$\frac{\sigma_X}{\sigma_Y} < 1$, X is more dependent on Y than Y dependent on X.
$\frac{\sigma_X}{\sigma_Y} > 1$, X is less dependent on Y than Y dependent on X.

V. Application of $o(X,Y)$ to Insurance Risk Modeling

We will apply the continuous $o(X,Y)$ measure to risk modeling for joint lives and survivor insurance and annuity contracts. The life expectancy of pairs of groups of insured persons is a pricing and reserving factor which must be considered when we want to assess the risks. Our general approach consists of the following steps:

1) Obtain by regression (optimization) the probability density functions from the available data samples;
2) Compute $o(X,Y)$ and $o(Y,X)$ using these density functions that passed the significance tests;
3) Analyze the unilateral interactions between variables X and Y.

If the regression fails from miscellaneous reasons or the model of the probability density function cannot be clearly selected, we can choose a non-parametric approach, e.g. a kNN based method, as presented in [7].

In our study, we have joint male-female pairs (married, but not necessarily from the same family), as data samples from the hypothized joint probability distribution (X,Y). Variable X is the number of deaths per thousand annual risk exposures of joint lives and survivors (males), given at least one death occurred in the group. Variable Y is the number of deaths per thousand of joint lives and survivors annual exposures (females), given that a death in that group occurred.

We use older age mortality rates by policy, derived from the database of the Individual Annuity Experience Committee (IAEC) of the Society of Actuaries with the contract data from major insurance companies based in the USA. We will compute our unilateral dependency measure to investigate the dependency of male mortality rate on female's for a general block of joint lives and survivors contracts. $o(X,Y)$ is not like most the traditional correlation but aims to quantify the unilateral causation relationship between X and Y. The causality relationship between X mortality rate (by policy) and one of the FARM (Factors Affecting Retirement Mortality) variables as Y may be measured and ranked for model factor selection, although it is outside the scope of this section. The unilateral dependency measure between random variables will be applied to mortality variable rate to hopefully aid stochastic model efficiency via incorporating mortality variable as a dynamic, experience-adjustable stochastic process, whose causation relationship should be recognized, evaluated, and quantified by a robust yet practical measure system. We will calculate $o(X,Y)$ directly from the joint probability model derived from a large multivariate mortality data sets. The final goal is to use our results for screening and selecting essential mortality model predictors from the FARM collection for the future research.

For the years 2000-2008, we have extracted 2276 subgroups of Joint & Survivor contracts, which are matched as (x,y) pairs belonging to the same risk group.

To compute the density functions of X, Y, joint (X,Y), and joint (Y,X) we first try to fit them into a joint probability density model, the first step when we want to apply our method. Since the true shape of the random variables X and Y is not known, we have to investigate the fitting of the data with various models and select the distribution which passes the statistical tests. The first assumption was they might be modeled by the normal distribution, but X and Y fit fails to pass the normal test. They are not normally distributed in probability. Thus, it is not practical to plug in a bivariate Gaussian model to calculate $o(X,Y)$.

However, they pass the lognormality test strongly, including for a bivariate lognormal distribution as assessed by applying the Kolmogorov-Smirnov and Anderson-Darling tests to the marginal distributions of X and Y. The testing of the bivariate joint density model of X and Y is assumed to pass via the well-known joint normal distribution, where joint normal is established if the marginal distribution is normal. This theorem is applied to the joint lognormal distribution in light of monotonic variable transformation. The resulted values are $o(X,Y) = 0.00519505$ and $o(Y,X) = 0.00529863$ which is an indicator that the group X (survival of males) has a slightly lower dependence on group Y (survival of females) than the vice versa [12].

For the second case study we refined the data and we set the variable X as the number of deaths per thousand age 65+ male joint lives and survivors, given at least a death occurred

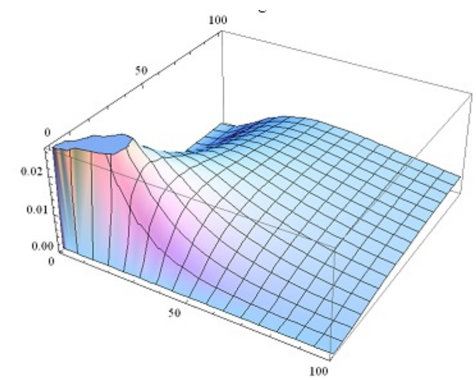

Fig. 3. Bivariate conditional lognormal probability density function of X given Y for the group of age 65+.

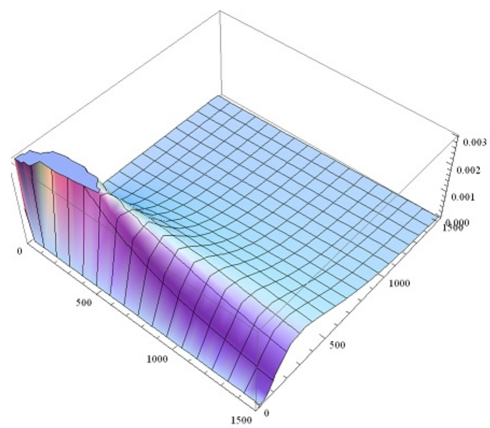

Fig. 4. Bivariate conditional lognormal probability density function of Y given X for the group of age 65+.

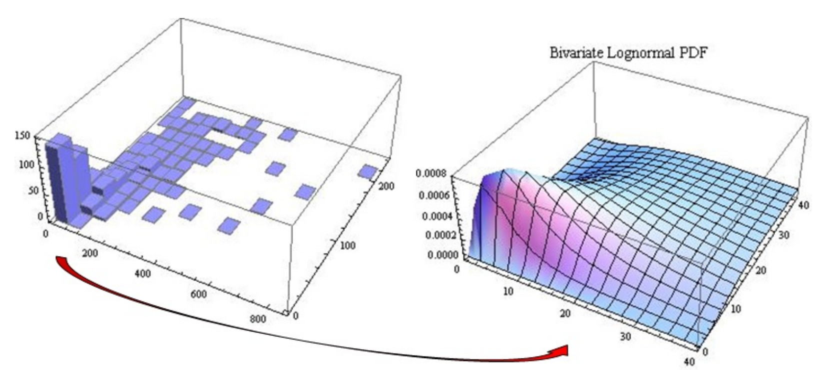

Fig. 5. 3D Histogram Based on 2000-2008 IAEC Database representing the intrinsic mortality model for all age joint lives. The random variable corresponds to the number of deaths per 1000 annual exposures. 583 older age groups of male-female pairs from 2276 joint lives groups, with zero truncated were used in this plot.

in all special groups and the variable Y as number of deaths per thousand age 65+ female joint lives and survivors, given at least a death occurred in all special groups. We found that both series also fit well with the lognormal distribution, as depicted in the 3D representations from Figs. 3 and 4. The obtained values are $o(X, Y) = 0.0042535$ and $o(Y, X) = 0.0043478$, showing once again that the group X (survival of males) has a slightly lower dependence on group Y (survival of females) than the vice versa.

When age independent, the mortality random routine for male or female groups all follow lognormal model very closely. The joint lives mortality seems to fit well by our bivariate lognormal model [12]. Although further statistical tests should be performed, the theory of bivariate normal model lends strong support of strong fit of a bivariate lognormal.

We conclude that the IAEC data consistently reveals $o(X, Y) < o(Y, X)$ to be considered by models.

This indicates an average of 5 out of 1000 group expe-

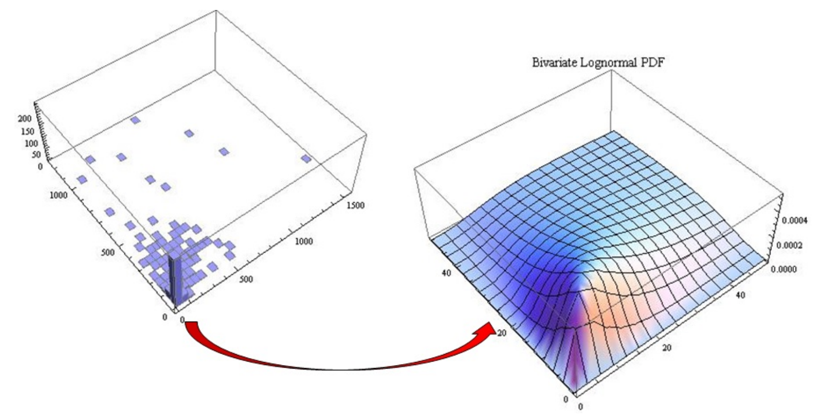

Fig. 6. 3D Histogram Based on 2000-2008 IAEC Database representing the intrinsic mortality model for older age (65+) joint lives. The random variable corresponds to the number of deaths per 1000 annual exposures. 523 older age groups of male-female pairs from 2276 joint lives groups, with zero truncated were used in this plot.

riences male's mortality rate depends on the female's for the policyholders contained in the 2000-2008 IAEC Database. For policyholders over 65 years old, the unilateral dependence is lower by one per thousand exposures, reducing to 4 cases out of 1000. As far as male's mortality rates depending on female's mortality rate more than the vice versa, the $o(X, Y)$ results support a tie, however, slightly leaning toward more female dependency on the male's mortality.

It will be interesting to see if such finding remains the same for uninsured populations in the U.S. for the future research, and also for the new data after 2008 which is becoming available in 2016, especially traditionally people may believe that the male mortality depends on their female partners.

VI. Conclusion

We have derived and proved some properties of the continuous unilateral dependence measure $o(X, Y)$ and relate them to Li's postulates. The following postulates are fulfilled: **P1**, **P2**, and **P4**. However, **P5** and **P6** are not satisfied. Postulate **P3** is only partially satisfied.

The $o(X, Y)$ measures the unilateral dependence between the two variables. The continuous version is very useful when we have to derive some mathematical properties regarding approximation techniques, as we did in [6] and [7]. The quite simple mathematical expression of $o(X, Y)$ makes this process easier than when using more complex nonsymmetric measures.

In the case studies, we showed that the higher volatility (largely scaled variable or larger standard deviation parameter, the random variable gives lower IE as in the Pareto example and lower $o(X, Y)$ as in the bivariate normal. This result is consistent with the concept of certainty measured by the IE and the $o(X, Y)$ measure.

The testing of bivariate lognormal density model of X and Y is assumed to pass via the well known joint normal distribution property where joint normal is established if marginal distribution is normal. This property is extended to lognormal model without mathematical proof included, but our intuition instead, supported by monotonic variable transformation.

With respect to our application area (mortality rate modeling for life insurance), the $o(X, Y)$ can be used to other intrinsic mortality models to help identify, investigate, analyze and model other inter-relationships among risk factors that drive the direction or magnitude of mortality rate for a specific insurance product, when the database is available and fit to parametric probability density function well.

In practice, insurance companies use deterministic mortality tables to price life insurance products and set the reserves to prevent insolvency. With increasing regulatory concern, there is a movement to stochastic models to better predict these risk factors, such as mortality rate, which is never a constant, but instead a variable subject to population change anti-selection and longevity. Finally, the $o(X, Y)$ may be need to to screen and rank important factors from the FARM variables to help improve actuarial risk models for older age retirement industry.

VII. Acknowledgments

The authors would like to acknowledge the Society of Actuaries for the IAEC Database, Jim Smigaj and Zachary Haberman for the AMOOF3 software, and Tom Edwalds for suggestions.

References

[1] Răzvan Andonie and Angel Caţaron. Energy generalized LVQ with relevance factors. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1421 – 1426 vol.2, july 2004.

[2] Răzvan Andonie and Angel Caţaron. An informational energy LVQ approach for feature ranking. In *European Symposium on Artificial Neural Networks 2004, pages In d-side publications*, pages 471–476, 2004.

[3] Răzvan Andonie and Angel Caţaron. Informational energy kernel for LVQ. In *Proceedings of the 15th international conference on Artificial neural networks: formal models and their applications - Volume Part II*, ICANN'05, pages 601–606, Berlin, Heidelberg, 2005. Springer-Verlag.

[4] Răzvan Andonie and Florian Petrescu. Interacting systems and informational energy. *Foundation of Control Engineering*, 11:53–59, 1986.

[5] C. B. Bell. Mutual information and maximal correlation as measures of dependence. *Ann. Math. Statist.*, 33(2):587–595, 06 1962.

[6] Angel Caţaron and Răzvan Andonie. Energy supervised relevance neural gas for feature ranking. *Neural Processing Letters*, 32(1):59–73, 2010.

[7] Angel Caţaron, Răzvan Andonie, and Yvonne Chueh. kNN estimation of the unilateral dependency measure between random variables. In *2014 IEEE Symposium on Computational Intelligence and Data Mining, (CIDM 2014), Orlando, FL, USA, December 9-12, 2014*, pages 471–478, 2014.

[8] Angel Caţaron, Răzvan Andonie, and Yvonne Chueh. Financial data analysis using the informational energy unilateral dependency measure. In *Proceedings of the International Joint Conference on Neural Networks, (IJCNN 2015), Killarney, Ireland, July 12-17, 2015*, 2015.

[9] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[10] Silviu Guiasu. *Information theory with applications*. McGraw Hill New York, 1977.

[11] Hui Li. On nonsymmetric nonparametric measures of dependence. arXiv:1502.03850, 2015.

[12] M.D. Mostafa and M.W. Mahmoud. On the problem of estimation for the bivariate lognormal distribution. *Oxford Journals*, 1964.

[13] Octav Onicescu. Theorie de l'information. energie informationelle. *C. R. Acad. Sci. Paris, Ser. A–B*, 263:841–842, 1966.

[14] J. C. Principe, D. Xu, and J. W. Fisher III. Information-theoretic learning. In Simon Haykin, editor, *Unsupervised Adaptive Filtering*, Wiley, New York, 2000.

[15] Alfréd Rényi. On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungarica*, 10:441–451, 1959.

[16] B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *Ann. Statist.*, 9:879–885, 1981.

[17] B.W. Silverman. *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1986.

[18] J. C. A. van der Lubbe, Y. Boxma, and D. E. Bockee. A generalized class of certainty and information measures. *Information Sciences*, 32(3):187–215, 1984.