Reduction to Independent Variables: From Normal Distribution to General Statistical Case to Fuzzy

Mourad Oussalah
Centre for Software Reliability
City University
Northampton Squares
London EC1V OHB UK
OMourad@csr.city.ac.uk

Hung T. Nguyen
Dept. Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

Vladik Kreinovich
Dept. Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu

Abstract

In many practical problems, we must combine ("fuse") data represented in different formats, e.g., statistical, fuzzy, etc. The simpler the data, the easier to combine them. Therefore, to combine complex data, it is desirable to "decompose" this complex data into simpler (easy-to-combine) data chunks.

It is well known that when we have n random variables $x_1,\ldots,x_n$ with a joint Gaussian distribution, then we can reduce them to n independent variables by an appropriate linear transformation

$$x_1,\ldots,x_n \to$$

$$y_1 = f_1(x_1,\ldots,x_n),\ldots,y_n = f_n(x_1,\ldots,x_n).$$

It is not so well known but also true that when we have $x_1,\ldots,x_n$ with a known joint probability distribution (not necessarily Gaussian), then we can always reduce them to n independent variables by an appropriate non-linear transformation. In this paper, we show that a similar result holds for fuzzy uncertainty as well.

Keywords: Reduction to independent variable, Probabilistic, Fuzzy.

1 Decomposition is Desirable for Data Fusion

In many practical problems, we must combine ("fuse") data represented in different formats, e.g., statistical, fuzzy, etc. The simpler the data, the easier to combine them. Therefore, to combine complex data, it is desirable to "decompose" this complex data into simpler (easy-to-combine) data chunks.

2 Known Fact: Gaussian Distributions of Several Variables Can Be Reduced to Independent Case

It is well known (see, e.g., [5], Section 4.8) that if we have a normal joint distribution of several random variables $x_1,\ldots,x_n$, then we can make an appropriate linear change of variables, from the original variables $x_1,\ldots,x_n$ to their appropriate linear combinations $y_1,\ldots,y_n$, after which the new random variables $y_1,\ldots,y_n$ are independent.

Since the random variables $y_1,\ldots,y_n$ are independent, their joint probability density function $\rho_y(y_1,\ldots,y_n)$ can be factorized, i.e., represented as a product

$$\rho_y(y_1,\ldots,y_n) =$$

$$\rho_1(y_1)\cdot\rho_2(y_2)\cdot\ldots\cdot\rho_n(y_n). \quad (1)$$

Moreover, it is known that we can choose a linear transformation $x_1,\ldots,x_n \to y_1,\ldots,y_n$ in such a way that each of the variables y_k depends only on the first k old variables $x_1,\ldots,x_k$. In other words, we can choose this transformation in the following way:

$$y_1 = a_{11} \cdot x_1;$$

$$y_2 = a_{21} \cdot x_1 + a_{22} \cdot x_2;$$

$$y_3 = a_{31} \cdot x_1 + a_{32} \cdot x_2 + a_{33} \cdot x_3;$$

$$\ldots$$

$$y_k = a_{k1} \cdot x_1 + a_{k2} \cdot x_2 + \ldots + a_{kk} \cdot x_k; \qquad (2)$$

$$\ldots$$

$$y_n = a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \ldots + a_{nk} \cdot x_k + \ldots + a_{nn} \cdot x_n,$$

where $a_{11}, a_{21}, \ldots$ are real numbers.

We can also define the transformation in such a way that not only the variables $y_1, \ldots, y_n$ are independent, but that each of them is distributed according to the standard Gaussian distribution, with 0 average and unit standard deviation.

3 Case of General Distributions (Not Necessarily Gaussian)

For a general (not necessarily Gaussian) distribution similar results hold; see, e.g., [1, 8].

First, if we have a joint distribution of several random variables $x_1, \ldots, x_n$ (for which the probability density is everywhere defined, so we do not have infinite density values), then we can make an appropriate non-linear change of variables, from $x_1, \ldots, x_n$ to $y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n)$, after which the new random variables $y_1, \ldots, y_n$ are independent.

Since the random variables $y_1, \ldots, y_n$ are independent, their joint probability density function $\rho_y(y_1, \ldots, y_n)$ can be factorized, i.e., represented as a product:

$$\rho_y(y_1, \ldots, y_n) =$$

$$\rho_1(y_1) \cdot \rho_2(y_2) \cdot \ldots \cdot \rho_n(y_n). \qquad (3)$$

It can also be proven that we can choose a transformation $x_1, \ldots, x_n \to y_1, \ldots, y_n$ in such a way that each of the variables y_k depends only on the first k old variables $x_1, \ldots, x_k$. In other words, we can choose this transformation in the following way:

$$y_1 = f_1(x_1);$$

$$y_2 = f_2(x_1, x_2);$$

$$y_3 = f_3(x_1, x_2, x_3);$$

$$\ldots$$

$$y_k = f_k(x_1, x_2, \ldots, x_k); \qquad (4)$$

$$\ldots$$

$$y_n = f_n(x_1, x_2, \ldots, x_k, \ldots, x_n),$$

where $f_1(x_1)$ is a function of one real variable, $f_2(x_1, x_2)$ is a function of two real variables, etc.

We can also define the transformation in such a way that not only the variables $y_1, \ldots, y_n$ are independent, but that each of them is uniformly distributed on the interval $[0, 1]$.

4 This Result Is Used in Computer Simulation of Random Variables

This is how this result is used in simulated different probability distributions. Let us assume that we are given the joint probability distribution, and we must simulate a random vector $x = (x_1, \ldots, x_n)$ which is distributed according to this distribution. To do that, we:

- find the above transformation (4) to independent uniformly distributed variables $y_1, \ldots, y_n$, and
- find the inverse transformation:

$$x_1 = g_1(y_1);$$

$$x_2 = g_2(y_1, y_2);$$

$$x_3 = g_3(y_1, y_2, y_3);$$

$$\ldots$$

$$x_k = g_k(y_1, y_2, \ldots, y_k); \qquad (5)$$

$$\ldots$$

$$x_n = g_n(y_1, y_2, \ldots, y_k, \ldots, y_n).$$

After that, to simulate the variables $x_1, \ldots, x_n$, we do the following:

- we first use standard random number generators to generate n independent random variables $y_1, \ldots, y_n$ which are uniformly distributed on the interval $[0, 1]$;
- then, we use the inverse transformation (2) to get the variables $x_1, \ldots, x_n$ with the desired distribution.

5 General Result Reformulated In More Abstract Mathematical Terms

This result can be reformulated in more abstract terms: that all non-atomic probability measures are isomorphic to each other in the sense that for every two of them, there is a measure-preserving transformation which leads from the first to the second one. This result was first proven in the 1940's.

- Such a transformation was explicitly used, e.g., by N. Wiener, when he first described Brownian motion (Wiener process) in precise probabilistic terms: for that description, he used an explicit Peano-like mapping from the set of all continuous functions to the interval $[0,1]$; see, e.g., [9, 10]. This transformation is also very useful in a constructive representation of Wiener processes; see, e.g., [3, 4].

- Recently, this result has been used to analyze the relation between Chu spaces and probabilities [6].

6 Idea of the Proof

6.1 1D Case

For a 1D case, when $n = 1$, the desired transformation is given by $y_1 = F(x_1)$, where $F(x)$ is a cumulative probability function of the probability distribution corresponding to x_1.

Indeed, by definition, for every real number a, the value $F(a)$ is equal to the probability $P(x_1 \le a)$ that the random variable x_1 is smaller than or equal to a. Since $F(x)$ is a monotonically increasing function, the inequality $x_1 \le a$ is equivalent to $F(x_1) \le F(a)$. Therefore,

$$P(F(x_1) \le F(a)) = P(x_1 \le a) = F(a). \qquad (6)$$

Thus, for $y_1 = F(x_1)$, we have $P(y_1 \le F(a)) = F(a)$ for every real number a. Since there is a density, the function $F(x_1)$ is differentiable hence continuous. Therefore, for every real number b from the interval $[0,1]$, there exists an a for which $b = F(a)$. Hence, for every real number b, $P(y_1 \le b) = b$. This means that y_1 is uniformly distributed on the interval $[0,1]$.

6.2 2D Case

In a 2D case, when $n = 2$ and we have two variables x_1 and x_2, we can:

- form a marginal distribution for x_1,
- get an appropriate transformation $x_1 \to y_1$, and
- then, for each value of x_1, apply a similar transformation $x_2 \to y_2$ (depending on x_1) to get a uniform distribution for y_2 too.

6.3 General Multi-D Case

Similarly, we can extend it to any number of variables $x_1, \ldots, x_n$.

7 In Fuzzy Case, A Similar Result is Also True

We can consider a similar problem for the fuzzy case:

- we have a membership function $\mu(x_1, \ldots, x_n)$ which is not necessarily representable as a product or a minimum of functions of one variable, and

- we would like to find a transformation to new variables $y_1, \ldots, y_n$ so that in the new variables, the membership function would become factorizable.

In fuzzy case, such a transformation is also always possible, and the proof of this possibility is even simpler than in the probabilistic case: namely, it is sufficient to take:

$$\begin{aligned} y_1 &= \mu(x_1, \ldots, x_n), \\ y_2 &= x_2, \\ y_3 &= x_3, \\ &\ldots, \\ y_k &= x_k, \\ &\ldots \\ y_n &= x_n. \end{aligned} \qquad (7)$$

Then, due to extension principle, for $y_1, \ldots, y_n$, the membership function is equal to

$$\mu_y(y_1, \ldots, y_n) = y_1. \qquad (8)$$

Since we can represent this membership function as

$$\mu_y(y_1,\ldots,y_n) =$$

$$\mu_1(y_1)\cdot\mu_2(y_2)\cdot\ldots\cdot\mu_n(y_n), \quad (9)$$

where $\mu_1(y_1) = y_1$ and $\mu_k(y_k) = 1$ for all $k > 1$, this membership function is clearly factorizable.

8 Reduction to Independent Variables: Case of min

In standard fuzzy logic, "and" can be represented both by the product and by the minimum; see, e.g., [11]. It therefore makes sense to look for min-decompositions. It turns out that the exact same membership functions provide a min-decomposition for the membership function $\mu_y(y_1,\ldots,y_n)$:

$$\mu_y(y_1,\ldots,y_n) =$$

$$\min(\mu_1(y_1),\mu_2(y_2),\ldots,\mu_n(y_n)). \quad (10)$$

9 Reduction to Independent Variables: Case of an Arbitrary T-Norm

More generally, for an arbitrary t-norm $f_\&(a,b)$ (see, e.g., [2, 7]), these same membership functions provide a decomposition corresponding to this t-norm:

$$\mu_y(y_1,\ldots,y_n) =$$

$$f_\&(\mu_1(y_1),\mu_2(y_2),\ldots,\mu_n(y_n)). \quad (11)$$

Acknowledgements

This work was supported in part by NASA grants NCC5-209 and NCC 2-1232, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-00-1-0365, by Grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund, and by NSF grants CDA-9522207 and 9710940 Mexico/Conacyt.

The authors are thankful to Lotfi A. Zadeh for the formulation of the problem and valuable discussions.

References

[1] A. H.-S. Ang and W. H. Tang (1990). *Probability Concepts in Engineering, Planning, and Design, Vol. II, Decision, Risk, and Reliability*, Wiley and Sons, New York.

[2] G. Klir and B. Yuan (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey.

[3] V. Kreinovich (1974). Applying Wiener measure to calculations, *Proceedings of the 3rd Leningrad Conference of Young Mathematicians*, Leningrad University, pp. 6–9 (in Russian).

[4] V. Kreinovich (1974a). Constructive theory of Wiener measure, *Proceedings of the 3rd USSR National Conference on Mathematical Logic*, Novosibirsk, USSR, 1974, pp. 116–118 (in Russian).

[5] A. Leon-Garcia (1989). *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, Massachusetts.

[6] H. T. Nguyen and N. Nguyen (2001). Chu spaces in uncertainty analysis, *Intern. Journal of Intelligent Systems*, Vol. 15, pp. 425–440.

[7] H. T. Nguyen and E. A. Walker (1999). *First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida.

[8] M. Rosenblatt (1952). Remarks on a multivariate transformation, *Annals of Math. Stat.*, 1952, Vol. 23, No. 3, pp. 470–472.

[9] N. Wiener (1961). *Cybernetics; or, Control and communication in the animal and the machine*, MIT Press, Cambridge, Massachusetts.

[10] N. Wiener (1964). *Extrapolation, interpolation, and smoothing of stationary time series, with engineering applications*, MIT Press, Cambridge, Massachusetts.

[11] L. A. Zadeh (1965). Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338–353.