

A Model of Fuzzy Normal Distribution

Haoge Liu, Jianhe Guan*

School of Science, China University of Geosciences (Beijing), Beijing, China

Email: 690594105@qq.com, *guanjh@cugb.edu.cn

How to cite this paper: Liu, H.G. and Guan, J.H. (2016) A Model of Fuzzy Normal Distribution. *Open Journal of Statistics*, 6, 749-755.

<http://dx.doi.org/10.4236/ojs.2016.65061>

Received: July 10, 2016

Accepted: September 19, 2016

Published: September 22, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Nowadays fuzzy concepts are frequently used as statistical parameters, while the traditional normal distribution can only accept determinate variable. In order to design a practical model for fuzzy statistic events, this paper combines the fuzzy number, like “may-occur”, “very-likely-occur”, “rarely-occur”, to optimize the normal distribution probability density function, to provide a significant method in statistics.

Keywords

Fuzzy, Normal Distribution, Composite Function’s Integral

1. Introduction

Normal distribution occupied an important position on probability theory and mathematical statistics both theoretically and practically. It widely exists in various fields including natural phenomenon, industrial production and high-technology, etc.

Generally speaking, normal distribution is an index influenced by many independent random factors, but each of them only has a tiny effect. For instance, quality indexes of all kinds of products (like size of tools, capacitance of capacitors, strength or tension of fibers), one characteristic of a group (like vital capacity of pupils in a class, plant length or stem diameter of rice in one area), or measured data of events (like the highest air temperature, average rainfall or humidity). All the indexes above are approximately obey the normal distribution.

However, as the development of science, especially the development of computer science and artificial intelligence, we are facing an important issue: how to use mathematical language to express human thinking. So after the birth of fuzzy mathematics in 1956, and a lot of combinations between fuzzy numbers and other mathematical theories were built rapidly. The fuzzy language is a method to transform natural language into quantification and mathematics, to adapt the needs of computer programs. The fuzziness is the essence of language, with which rich emotion can be expressed. In

mathematical statistics, “may-occur”, “very-likely-occur” or “rarely-occur” are vague expressions, but we can use the fuzzy membership function to quantify, and fix them into the calculation of the normal distribution.

CM Stein published in 1981 that using sum of squared errors as loss, estimation of the ways of normal in dependent random variables is advised, then he suggested that mean vector centered at an arbitrary estimate could be an application of calculating approximate sets [1]. In 2007, Mikl applied normal distribution for statistics on the Stirling Permutation defined by Gessel and Stanley, he equidistributed statistics on these objects converge to a normal distribution [2]. Even recently, CFF Karney published an algorithm for sampling from the normal distribution exactly, whose parameters are rational numbers, thereafter uniform random digit are copied into the representation directly [3]. Ever since the built of fuzzy maths, it’s application with calculation algorithm has been prospered [4]-[9].

Traditional normal distribution can only calculate the determined indexes, such as “40 degrees Celsius” or “120 pounds” or “45 years old”. But the uncertain concepts of data which can be represented by fuzzy number, like “appropriate temperature”, “normal weight” or “middle-aged”, also need to be count. So in this paper, we provide a model to calculate the fuzzy concepts’ probability by the integral of normal distribution probability density function.

2. Fuzz Formula

The traditional normal distribution formula is showed below:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

In this formula, the “ μ ” is defined as the expectation of a random event, and the “ σ ” is defined as the variance of a random event, they are identified when dealing with a determinate question. Usually the “ $x = g$ ” should be a specific number when we check a certain concept, and $f(x)$ can be calculated by determinate automaton but the “ g ” sometimes can be calculated as a fuzzy number. Fuzzy numbers are discussed in the interval number’s basis, the interval number is considering a closed interval $[A, B]$ as a number of processing, Each independent variable x in $[A, B]$ is mapped to the corresponding normal distribution density value, and at the same time, it also satisfies the corresponding membership function. So the integral of $f(x) * g(x), x \in [A, B]$ is the probability of fuzzy number.

We set $H(x)$ as the probability value of fuzzy number “ g ”, so:

$$H(x) = \int_{-\infty}^{+\infty} h(x) dx = \int_{-\infty}^{+\infty} f(x) * g(x) dx = \int_A^B f(x) * g(x) dx$$

And this paper gives three examples:

2.1. May-Occur Events

As the standard normal distribution table, the integral quantity from $\varphi(-0.5)$ to $\varphi(0.5)$ is 0.383, and the value from $\varphi(-1)$ to $\varphi(1)$ is 0.6826 (see **Figure 1**).

So the possibility between $\Phi(\mu - \sigma)$ and $\Phi(\mu + \sigma)$ occupy mostly, and the events $[\mu - \sigma/2, \mu + \sigma/2]$ can belong to the “may-occur event”. In this paper, we defined the result of membership function between $[\mu - \sigma/2, \mu + \sigma/2]$ is 1, And the result of membership function between $\Phi(\mu - \sigma)$ and $\Phi(\mu - \sigma/2)$ as well as events between $\Phi(\mu + \sigma/2)$ and $\Phi(\mu + \sigma)$ are lifting from zero as it approach to the middle, so according to the knowledge in fuzzy maths we defined it changes as quadratic function (see **Figure 2**).

So the membership function $g(x)$ should be divided into five parts and here we gives the may-occur events’ membership function:

$$g(x) = \begin{cases} 0, & x \leq \mu - \sigma \\ \left(\frac{x - \mu + \sigma}{\sigma/2}\right)^2, & \mu - \sigma < x < \mu - \sigma/2 \\ 1, & \mu - \sigma/2 \leq x \leq \mu + \sigma/2 \\ \left(\frac{\mu + \sigma - x}{\sigma/2}\right)^2, & \mu + \sigma/2 < x < \mu + \sigma \\ 0, & \mu + \sigma \leq x \end{cases}$$

And its chant is shown in **Figure 3**.

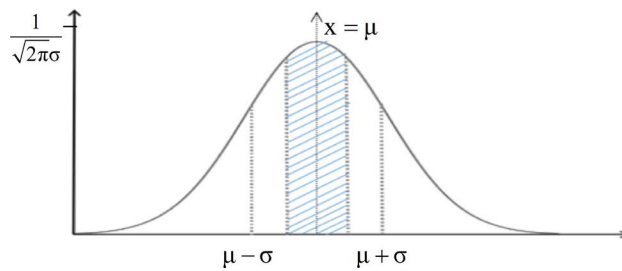


Figure 1. The possibility of Normal distribution from $\Phi(\mu - \sigma/2)$ to $\Phi(\mu + \sigma/2)$.

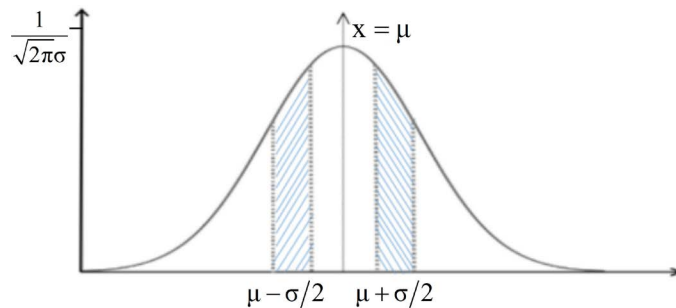


Figure 2. The possibility of Normal distribution from $\Phi(\mu - \sigma)$ to $\Phi(\mu - \sigma/2)$ and from $\Phi(\mu + \sigma/2)$ to $\Phi(\mu + \sigma)$.

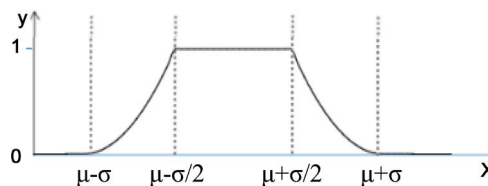


Figure 3. The membership function of may-occur event.

This kind of model have applications in multiple ways, some broad-loose events like “general quality products in one production” or “normal section of score in a test” which may happen most time we sample, is a result of general research expression.

2.2. Very-Likely-Occur Event

As the standard normal distribution table, the integral quantity from $\varphi(-0.2)$ to $\varphi(0.2)$ is 0.1586.

So the event between $[\mu - \sigma/5, \mu + \sigma/5]$ can be called “very-likely-occur event” cause they are surrounding the symmetry axis “ $x = \mu$ ”, and the membership function in this interval would be 1. The interval $(\mu - \sigma, \mu - \sigma/5)$ and from $(\mu + \sigma/5, \mu + \sigma)$ also agree with the power function relationship, but its tendency will change more steeply, we define it as cubic function.

So the very-likely-occur events’ membership function shows below:

$$g(x) = \begin{cases} 0, & x \leq \mu - \sigma \\ \left(\frac{x - \mu + \sigma}{4\sigma/5}\right)^3, & \mu - \sigma < x < \mu - \sigma/5 \\ 1, & \mu - \sigma/5 \leq x \leq \mu + \sigma/5 \\ \left(\frac{\mu + \sigma - x}{4\sigma/5}\right)^3, & \mu + \sigma/5 < x < \mu + \sigma \\ 0, & \mu + \sigma \leq x \end{cases}$$

And its chart is shown in **Figure 4**.

This kind of model have applications in every investigation, the most frequent events like “the time office workers usually get up” or “the speed of one train in one trip” which may happen 90 percent each time we sample, it’s basically distribute around a certain point.

2.3. Rarely-Occur Event

As the standard normal distribution table, the integral quantity from $\varphi(-3)$ to $\varphi(3)$ is 0.9974.

So the event between $(-\infty, \mu - 3\sigma]$ and $[\mu + 3\sigma, +\infty)$ can be called “rarely-occur event” cause their probability are almost zero. And member function in $[\mu - 2\sigma, \mu + 2\sigma]$ is 0.

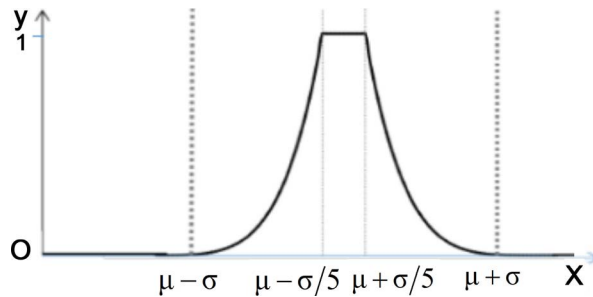


Figure 4. The membership function of very-likely-occur event.

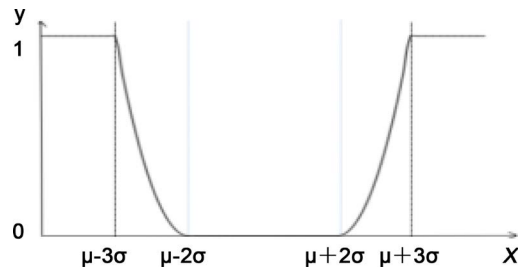


Figure 5. The membership function of rarely-occur event.

So the rarely-occur events' membership function shows:

$$g(x) = \begin{cases} 1, & x \leq \mu - 3\sigma \\ \left(\frac{x - \mu + 2\sigma}{\sigma}\right)^2, & \mu - 3\sigma < x < \mu - 2\sigma \\ 0, & \mu - 2\sigma \leq x \leq \mu + 2\sigma \\ \left(\frac{x - \mu - 2\sigma}{\sigma}\right)^2, & \mu + 2\sigma < x < \mu + 3\sigma \\ 1, & \mu + 3\sigma \leq x \end{cases}$$

And its chart is shown in **Figure 5**.

And this is the model of extreme situation.

3. Case Analysis

In a sports comprehensive evaluation, a class of 50 people's score shows as follows:

53.6162 55.6079 61.4699 61.9634 64.7698 65.8982 66.3626 66.8872
 70.7759 71.1514 72.1262 72.3638 73.5902 74.0450 74.1331 74.1602
 75.5567 76.7973 78.0933 78.5431 78.9637 79.1533 80.3374 80.5220
 80.5663 80.6490 80.6914 81.0579 81.2104 81.6568 81.7275 82.1617
 82.2923 83.6559 83.6559 84.6994 85.2302 85.4510 85.7813 86.4338
 86.8235 87.2359 88.7201 88.9829 89.6834 90.4821 90.7801 91.4503
 93.5038 99.7933

Fitting these data. The results can be regarded as a normal distribution, with $f(x)$ to represent normal distribution function, $g(x)$ represent the membership functions, the mathematical expectation is approximately 80, the standard deviation of approximately 10, so according to the membership function, $\mu - \sigma = 70$, $\mu + \sigma = 90$, $\mu - \sigma/2 = 75$, $\mu + \sigma/2 = 85$.

1) To get the probability of "normal level" score should based on the "may-occur" model, the operation procedure shows as follows:

$$h(x) = f(x) * g(x) = \begin{cases} 0, & x \leq 70 \\ f(x) * \left(\frac{x - 70}{5}\right)^2, & 70 < x < 75 \\ f(x), & 75 \leq x \leq 85 \\ f(x) * \left(\frac{90 - x}{5}\right)^2, & 85 < x < 90 \\ 0, & 90 \leq x \end{cases}$$

$$\begin{aligned}
 H(x) &= \int_{-\infty}^{+\infty} h(x) dx \\
 &= \int_{70}^{75} f(x) * ((x-70)/5)^2 dx + \int_{75}^{85} f(x) dx + \int_{85}^{90} f(x) * ((90-x)/5)^2 dx \\
 &= 0.0545 + 0.3829 + 0.0545 = 0.492
 \end{aligned}$$

In this case, “may-occur” events’ probability is 0.492.

2) To get the probability of “very poor level” score should based on the “may-occur” model, and we only use half model which less than expectation, the operation procedure shows as follows:

$$p(x) = f(x) * g(x) = \begin{cases} 1, & x \leq 50 \\ f(x) * ((x-60)/10)^2, & 50 < x < 60 \\ 0, & 60 \leq x \end{cases}$$

$$\begin{aligned}
 P(x) &= \int_{-\infty}^{+\infty} p(x) dx = \int_{-\infty}^{50} f(x) dx + \int_{50}^{60} f(x) * ((x-60)/10)^2 dx \\
 &= 0.0013 + 0.0035 = 0.0048
 \end{aligned}$$

In this case, “rarely-occur” events’ probability is 0.0048.

These two case also conform to reality. And for all those three models, the result will be changed if μ or σ were recalculated. So if we remove the intervals of event, we will get a different possibility.

4. Conclusions

In this paper, we design a calculation model to fuzz normal distribution probability density function, so that the normal distribution can be closer to the actual life. In this paper, three distributed events “may-occur”, “very-likely-occur” and “rarely-occur” were provided, and the results in tests also turn out to be logical. But as the model designed, we couldn’t give more references to verify the effectiveness compared to other fuzzy membership functions, like linear function. We can only say that each of those viewpoints has some reason for now, and our future work will be the comparison of their advantages and disadvantages.

If other distributed events are all provided, there may be a general model be created to define them. So if we can refine human emotion more specific and meticulous, this model will have a very wide range of applications in the field of artificial intelligence.

References

- [1] Stein, C.M. (1981) Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics*, **9**, 1135-1151.
- [2] Bona, M. (2007) Real Zeros and Normal Distribution for Statistics on Stirling Permutations Defined by Gessel and Stanley. *SIAM Journal on Discrete Mathematics*, **23**, 401-406.
- [3] Karney, C.F.F. (2016) Sampling Exactly from the Normal Distribution. *ACM Transactions on Mathematical Software*, **42**, Article No. 3.
- [4] Astesiano, E. (Ed.) (1998) In: *Fundamental Approaches to Software Engineering (FASE98)*, *Lecture Notes in Computer Science*, Vol. 1382, Springer, Berlin.
- [5] Bartlett, K.A., Scantlebury, R.A. and Wilkinson, P.T. (1969) A Note on Reliable Full-Duplex

- Transmission over Half-Duplex Links. *Communications of the ACM*, **12**, 260-261.
<http://dx.doi.org/10.1145/362946.362970>
- [6] Eilouti, B.H. (2008) A Formal Language for Palladian Palazzo Façades Represented by a String Recognition Device. *Nexus Network Journal*, **10**, 245-267.
- [7] Araki, K., Galloway, A. and Taguchi, K. (Eds.) (1999) *1st International Conference on Integrated Formal Methods*, Springer, Berlin, June 1999.
- [8] Ahmed, E. (1996) Fuzzy Cellular Automata Models in Immunology. *Journal of Statistical Physics*, **85**, 291-294.
- [9] Wu, Q.-E., Pang, X.-M. and Han, Z.-Y. (2011) Fuzzy Automata System with Application to Target Recognition Based on Image Processing. *Computer and Mathematics with Applications*, **61**, 1267-1277. <http://dx.doi.org/10.1016/j.camwa.2010.08.101>



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojs@scirp.org