6 Sigma Rules Under Uncertain Normal Distributions

Danni Guo¹ and Renkuan Guo²
¹Kirstenbosch Research Center, South African National Biodiversity Institute, Private Bag X7, Claremont 7735, Cape Town, South Africa
²Department of Statistical Sciences, University of Cape Town, Private Bag, Rondebosch, 7701, Cape Town, South Africa

Abstract: In evaluating modeling quality, risk level and general quality management "6 sigma rules" are popular among statisticians and quality engineers today. The "6 sigma rules" of Gaussian distribution are very simple and elemental. If switching the working environment from the probability measure based to the uncertain measure based, the simple "6 sigma rules" will be no longer simple. In this paper, we investigate the problems when facing Liu's uncertain normal distribution which can only facilitate interval variance or standard deviation. Consequently, the way to define "uncertain 6 sigma rules" are described and discussed and thus those to practical applications.

Keywords: uncertain measure, uncertain normal distribution, Gaussian distribution, 6 sigma rules

I. Introduction

In this paper, we will investigate "6 sigma rules" under uncertain normal distributions. The meaning of the 6 sigma rules is not in the sense as quality management strategy "6 sigma" initiated by Motorola in 1986 [7]. Rather, we use term "6 sigma rules" to refer to as an elemental fact of a Gaussian (Normal) distribution. In probability theory [1] the Gaussian (Normal) distribution plays a fundamental role and facilitates the statistical modeling, estimation and inference. Practically no matter engineers in industries or statisticians in various fields rely on the knowledge on mean and standard deviation because the risk level and quality are quantified by these two parameters.

The way to specify "6 sigma rules" may be a too easy and elemental to be ignored by statisticians. We should be fully aware that the simplest "6 sigma rules" are composed of the foundation on which the modern quality control and Markov pattern analysis are rooted.

Figure 1 demonstrates the rules (i.e., the probability laws actually [9]) that the probability grades fall between unit sigma $\sigma$ away from mean $\mu$. Table I further emphasizes the probability rules that Figure 1 shown.

An example is the strange quality problems occurred in Japanese car industry. If the traditional probabilistic "6 sigma rules" really exist objectively, then the out control (beyond 6 sigma) is 0.27% level. But what happened from time to time, the out control (beyond 6 sigma) risk level is much higher than that of probability law. These facts may imply that an industrial environment may be explained by Liu's uncertain measure theory [2, 3, 4, 5, 6, 7] because of the production system complexity and human behavior involvement.

However, what we are interested to ask is: does uncertain theory have the counterpart - "the uncertain 6 sigma rules"? To address such a simple question, we have to review the relevant uncertain variable and distribution theory.

It is well known that the probability distribution upon which a random variable is fully defined, is equivalent to corresponding probability measure [1]. But in uncertain theory [2, 3, 4, 5, 6, 7], an uncertain distribution cannot fully specify an uncertain variable because an uncertain variable can only completely defined by an uncertain measure!

The structure of the remaining sections is stated as follows. Section II will be used to review Liu's axiomatic uncertain measure and uncertain normal variable theory [7]. In Section III, we will derive the interval variance given

**Table I: Area's probability grades between sigma increments (standard deviations)**

<table>
<thead>
<tr>
<th>Area between the interval limits</th>
<th>Probability grades of a (standard) Gaussian (normal) distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(-\infty,-3]$</td>
<td>0.00135</td>
</tr>
<tr>
<td>[-3,-2]</td>
<td>0.02140</td>
</tr>
<tr>
<td>[-2,-1]</td>
<td>0.13591</td>
</tr>
<tr>
<td>[-1,0]</td>
<td>0.34134</td>
</tr>
<tr>
<td>[0,1]</td>
<td>0.34134</td>
</tr>
<tr>
<td>[1,2]</td>
<td>0.13591</td>
</tr>
<tr>
<td>[2,3]</td>
<td>0.02140</td>
</tr>
<tr>
<td>[3,+$\infty)$</td>
<td>0.00135</td>
</tr>
</tbody>
</table>

Figure 1: Area’s percentage (measuring grades) between sigma increments (standard deviations) under a Gaussian (normal density) curve
Liu’s ([7]) uncertain normal distribution. Section IV serves the discussions on the uncertain 6 sigma rules given Liu’s ([7]) uncertain normal distribution. Section V concludes the paper.

II. Uncertain Normal Variable

Since an uncertain variable is fully defined by its uncertain measure, it is necessary to review the uncertain measure theory. Uncertain measure is an axiomatically defined set function mapping from a $\sigma$-algebra of a given space (set) to the unit interval $[0,1]$, which provides a measuring grade system of an uncertain phenomenon and facilitates the formal definition of an uncertain variable.

Let $\Xi$ be a nonempty set (space), and $\mathfrak{A}(\Xi)$ the $\sigma$-algebra on $\Xi$. Each element, let us say, $A \subset \Xi, A \in \mathfrak{A}(\Xi)$ is called an uncertain event. A number denoted as $\lambda(A), 0 \leq \lambda(A) \leq 1$, is assigned to event $A \in \mathfrak{A}(\Xi)$, which indicates the uncertain measuring grade with which event $A \in \mathfrak{A}(\Xi)$ occurs. The normal set function $\lambda(A)$ satisfies following axioms given by Liu [7]:

**Axiom 1:** (Normality) $\lambda(\Xi) = 1$.

**Axiom 2:** (Self-Duality) $\lambda(\emptyset) = 0, \lambda(A) + \lambda(A^c) = 1$.

**Axiom 3:** (Subadditivity) $\lambda\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \lambda(A_i)$ for any countable event sequence $\{A_i\}$.

**Definition 2.1:** (Liu [6, 7]) Any set function $\lambda: \mathfrak{A}(\Xi) \rightarrow [0,1]$ satisfies Axioms 1-3 is called an uncertain measure. The triple $\Xi, \mathfrak{A}(\Xi), \lambda$ is called the uncertain measure space.

**Definition 2.2:** (Liu [6, 7]) An uncertain variable $\xi$ is a measurable mapping, i.e., $\xi: (\Xi, \mathfrak{A}(\Xi)) \rightarrow ([0,1], \mathfrak{B}(\mathbb{R}))$, where $\mathfrak{B}(\mathbb{R})$ denotes the Borel $\sigma$-algebra on $\mathbb{R} = (-\infty, +\infty)$.

**Definition 2.3:** (Liu [6, 7]) The uncertain distribution $\Lambda: \mathbb{R} \rightarrow [0,1]$ of an uncertain variable $\xi$ on $\Xi, \mathfrak{A}(\Xi), \lambda$ is

$$\Lambda(x) = \lambda\{\tau \in \Xi | \xi(\tau) \leq x\}$$

(1)

An uncertain variable is completely specified by its uncertain measure, while an uncertain variable is only partially defined by the corresponding uncertain distribution. In other words, using an uncertain distribution to specify an uncertain variable, certain uncertainty will be brought in.

Liu [2, 3, 4, 5, 6, 7] proposed his version of normal uncertain variable with the uncertain distribution different from that in probability theory. Let us review Liu’s [4] normal uncertain variable and its distribution.

**Definition 2.4:** (Liu [6, 7]) An uncertain variable $\xi$ is called normal if its uncertain distribution takes the form

$$\Lambda(x) = \frac{1}{1 + e^{-\frac{2\sigma^2}{\sqrt{3}\lambda}}}}, \quad x \in \mathbb{R}$$

(2)

**Definition 2.5:** (Liu [6, 7]) An uncertain normal distribution is standard if the uncertain distribution takes the form

$$\Lambda_{0}(x) = \frac{1}{1 + e^{-\frac{x}{\sqrt{3}\lambda}}}}, \quad x \in \mathbb{R}$$

(3)

The standard uncertain normal distribution function ($\mu = 0, \sigma = 1$) is plotted in Figure 2:

![Figure 2](image-url)

Figure 2: The standard uncertain normal distribution ($\mu = 0, \sigma = 1$).

**Remark 2.6:** The normal uncertain distribution has a derivative function

$$\lambda(x) = \frac{\pi}{2\sqrt{3}\lambda} \left(1 + \text{ch}\left(\frac{\pi}{\sqrt{3}\lambda}(x - \mu)\right)\right), \quad x \in \mathbb{R}$$

(4)

We can establish that there is difference in areas under the Gaussian density curve and Liu’s normal derivative function curve. Table II lists the area’s measuring grades between unit increments under the uncertain standard normal derivative function curve. It should be aware that we call $\lambda(x)$ the derivative function of an uncertain normal distribution function, whose density function is undefined.

Table II: Area’s measuring grades between unit increments under the standard normal derivative function
Figure 3: Area’s measuring grades between unit increments (standard deviations) under an uncertain normal derivative function curve.

Figure 3 demonstrates the area’s measuring grades between unit increments under the standard normal derivative function curve. Immediately, we see some similarity between Figure 1 and Figure 3, but we have to emphasize the terminology difference because of the different measure systems. We utilize Table III to summarize the similarity between Liu’s normal uncertain variable (standard and general) and Gaussian (normal) random variable (standard and general).

TABLE III: Basic comparisons between Liu’s normal uncertain variable and Gaussian (i.e., normal) random variable.

<table>
<thead>
<tr>
<th>Area’s grades between the interval limits under derivative curve</th>
<th>Uncertain standard normal uncertain distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(-\infty, -3]$</td>
<td>0.004315</td>
</tr>
<tr>
<td>[-3, -2]</td>
<td>0.021577</td>
</tr>
<tr>
<td>[-2, -1]</td>
<td>0.114288</td>
</tr>
<tr>
<td>[0, 1]</td>
<td>0.359820</td>
</tr>
<tr>
<td>[1, 2]</td>
<td>0.114288</td>
</tr>
<tr>
<td>[2, 3]</td>
<td>0.021577</td>
</tr>
<tr>
<td>$[3, +\infty)$</td>
<td>0.004315</td>
</tr>
</tbody>
</table>

Up to now, what we see is the high similarity between Liu’s normal uncertain variable and Gaussian (i.e., normal) random variable. Does the similarity lead us to have “6 sigma rules” in applying uncertain normal variable? Section III will give further exposure.

III. Variance of An Uncertain Normal Distribution

Given Distribution Function

Let us start with the general definition of the expectation of an uncertain variable, denoted by $\mu = E[\xi]$.

**Definition III.1:** (Liu, [6], [7]) Let $\xi$ be a uncertainty variable on an uncertainty measure space $(\Xi, \mathcal{A}(\Xi), \lambda)$. The expectation $\xi$ is defined by

$$E[\xi] = \int_{-\infty}^{\infty} \lambda(\{\xi \geq r\}) dr - \int_{-\infty}^{0} \lambda(\{\xi \leq r\}) dr$$

provided that one of the two integrals exists at least.

**Theorem III.2:** (Liu, [7]) Let $\xi$ be an uncertain variable with a given uncertain distribution function $\Psi$, having a finite expectation $\mu$. Then the upper bound of the variance of $\xi$ is

$$\sigma_{\xi}^2 = 2 \int_{0}^{\infty} \left( r(1 - \Psi_{\xi}(r + \mu)) + (\Psi_{\xi}(\mu - r)) \right) dr$$

**Proof:**
\[ E[(\xi - \mu)^2] = \int_{-\infty}^{\infty} x(\xi - \mu)^2 \leq x \, dx = \int_{0}^{\infty} \lambda(\xi \geq \mu + \sqrt{x}) + \lambda(\xi \geq \mu - \sqrt{x}) \, dx \]

By changing variable \( \sqrt{x} = r \).

**Theorem III.3:** (Liu, [7]) Let \( \xi \) be an uncertain variable with a given uncertain distribution function \( \Psi_\xi \) having a finite expectation \( \mu \). Then the lower bound of the variance of \( \xi \) is

\[ \sigma_{\xi, l}^2 = \int_{0}^{\infty} (1 - \Psi_\xi(\mu + \sqrt{x})) \, dx \vee \int_{0}^{\infty} \Psi_\xi(\mu - \sqrt{x}) \, dx \quad (8) \]

**Proof:** Based on Theorem III.3 and Theorem III.4, the variance of an uncertain normal variable takes an interval value:

\[ \sigma_{\xi, l}^2 \leq V[\xi] \leq \sigma_{\xi, u}^2 \quad (12) \]

Examining the ratio of \( \sigma_{\xi, l}^2 / \sigma_{\xi, u}^2 \) under given uncertain normal distribution function in Eq. (10):

\[ \frac{\sigma_{\xi, l}^2}{\sigma_{\xi, u}^2} = \frac{\int (1 - \Lambda(\mu + \sqrt{x})) \, dx \vee \int \Lambda(\mu - \sqrt{x}) \, dx}{\int (1 - \Lambda(\mu + \sqrt{x})) \, dx + \int \Lambda(\mu - \sqrt{x}) \, dx} \]

\[ = \frac{1}{2} \]

The ratio is 1/2, which leads to the conclusion of the theorem.

**Remark III.5:** In Liu's work, the variance of an uncertain normal variable taking the distribution function as shown in Eq. (10) is stipulated as upper bound and denoted as \( \sigma^2 \). It is definitely an advantage in theoretical developments to have the stipulation. The parameter \( \sigma^2 \) has direct link with variance.

\[ V[\xi] = E[(\xi - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \left( \frac{1}{1 + e^{-\frac{x - \mu}{\sqrt{\sigma^2}}} \right) \, dx = \frac{3\sigma^2}{\pi^2} \int_{0}^{1} \left( \ln u - \ln(1 - u) \right)^2 \, du \]

\[ = \frac{3\sigma^2}{\pi^2} \left( 4 - 2 \times \left( 2 - \frac{\pi^2}{6} \right) \right) = \sigma^2 \]

Nevertheless, the interval form of a variance should deserve further attention because there is no counterpart in probabilistic statistics, in which a statistic is supposed be a scalar real-valued function.

**IV. Uncertain 6 Sigma Rules**

R.E. Moore [8] proposed interval algorithm theory in 1965. It involves inevitable the interval uncertainty concept. We could not find a very strict definition of it, rather, we would like interpret it intuitively. For example, \( V[\xi] \) is interval-valued, which can be expressed as \([\sigma_{\xi, l}^2, \sigma_{\xi, u}^2]\). The interval
uncertainty means that any value between the lower bound and the upper bound of the interval variance could be the variance with certain risk. In other words, the lower bound \( \sigma^2_{\xi, a} \) could be the variance, or, the upper bound \( \sigma^2_{\xi, u} \) could be the variance too. Actually, \( \forall \sigma^2_{\xi, u} \in [\sigma^2_{\xi, a}, \sigma^2_{\xi, u}] \) could be defined as the variance of uncertain variable \( \xi \) with certain degree of risk.

**Theorem IV.1:** Let an uncertain variable \( \xi \) defined by its uncertain distribution having two parameters. If the lower bound over the upper bound ratio of the interval variance \([\sigma^2_{\xi, a}, \sigma^2_{\xi, u}]\), \( \sigma^2_{\xi, a}/\sigma^2_{\xi, u} = 0.5 \), then the uncertain distribution takes a form:

\[
\Lambda(x) = \frac{1}{1 + e^{-\frac{1}{\sqrt{2}}(x - \mu)}} \text{, } x \in \mathbb{R}. \tag{15}
\]

where \( \mu \) is the expectation and \( \sigma^2 \) is the stipulated variance parameter.

Theorem IV.1 is a converse of Theorem III.4. As a matter of fact, among two parametric distribution family, the lower bound over the upper bound ratio of the interval variance \([\sigma^2_{\xi, a}, \sigma^2_{\xi, u}]\) reaches the minimum, 1/2. Therefore, the ratio of the two bounds indicates the uncertain distribution.

When discussing the uncertain 6 sigma rules, it is inevitable to engage interval-valued standard deviation.

**Definition IV.2:** Let the variance of uncertain variable \( \xi \) be \([\sigma^2_{\xi, a}, \sigma^2_{\xi, u}]\). Then the standard deviation interval is defined by

\[
\sigma_{\xi, \text{Std}} \triangleq \left[ \sqrt{\sigma^2_{\xi, a}}, \sqrt{\sigma^2_{\xi, u}} \right] = [\sigma^2_{\xi, a}, \sigma^2_{\xi, u}]. \tag{16}
\]

**Theorem IV.3:** Given an uncertain normal distribution with a form as Eq.(15), the standard deviation interval \( \sigma_{\xi, \text{Std}} \) can be expressed by \( \left[ \sqrt{\sigma^2_{\xi, a}}, \sqrt{\sigma^2_{\xi, u}} \right] \).

Therefore, the discussions of the uncertain 6 sigma rules rely on the specification of the sigma which is the "unit". We calculate the sizes of measure grade under three different sigma units the lower bound \( \sigma_{\xi, a}/\sqrt{2} \), the midpoint \( (1+1/\sqrt{2})\sigma_{\xi, a}/2 \), and the upper bound \( \sigma_{\xi, u} \) for a given uncertain normal distribution curve. Table III lists those values.

<table>
<thead>
<tr>
<th>Area coverage</th>
<th>Lower bound ( \sigma_{\xi, a}/\sqrt{2} )</th>
<th>Midpoint ( (1+1/\sqrt{2})\sigma_{\xi, a}/2 )</th>
<th>Upper bound ( \sigma_{\xi, u} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((-\infty, -3\sigma])</td>
<td>0.020884</td>
<td>0.009523</td>
<td>0.004315</td>
</tr>
<tr>
<td>((-3\sigma, -2\sigma])</td>
<td>0.050534</td>
<td>0.033736</td>
<td>0.021577</td>
</tr>
<tr>
<td>((-2\sigma, -\sigma])</td>
<td>0.145698</td>
<td>0.132092</td>
<td>0.114288</td>
</tr>
<tr>
<td>((-\sigma, 0])</td>
<td>0.282883</td>
<td>0.324650</td>
<td>0.359820</td>
</tr>
<tr>
<td>((0, \sigma])</td>
<td>0.282883</td>
<td>0.324650</td>
<td>0.359820</td>
</tr>
<tr>
<td>((\sigma, 2\sigma])</td>
<td>0.145698</td>
<td>0.132092</td>
<td>0.114288</td>
</tr>
<tr>
<td>((2\sigma, 3\sigma])</td>
<td>0.050534</td>
<td>0.033736</td>
<td>0.021577</td>
</tr>
<tr>
<td>([3\sigma, +\infty))</td>
<td>0.020884</td>
<td>0.009523</td>
<td>0.004315</td>
</tr>
</tbody>
</table>

From Table III, it can claim that the stipulated standard deviation \( \sigma = \sigma_{\xi, a} \) as sigma "unit" will give the highest coverage between -3sigma and +3sigma interval, 0.991371, while beyond \( \pm 3 \) sigma intervals, the coverage is 0.008629.

If we takes the standard deviation \( \sigma = \sigma_{\xi, a}/\sqrt{2} \) as sigma unit, coverage between -3sigma and +3sigma interval, 0.958231, while beyond \( \pm 3 \) sigma intervals, the coverage is 0.041769, which is 4.84 times larger than that of the stipulated sigma unit. If we takes the standard deviation \( \sigma = \sigma_{\xi, a}/\sqrt{2} \) as sigma unit, coverage between -3sigma and +3sigma interval, 0.980955, while beyond \( \pm 3 \) sigma intervals, the coverage is 0.019045, which is 2.21 times larger than that of the stipulated sigma unit.

**Remark IV.4:** Because most of the statisticians and engineers are familiar with probabilistic 6 sigma rules, sometimes it is unconscious to use probabilistic 6 sigma rules into the environment under uncertain distribution. Nevertheless, a wrong choice of sigma unit could lead to a 4.84 times of higher risk level! Therefore, the identification of working environment is critical: is it governed by probability measure? Or, is it governed by the uncertain measure.

V. Concluding Remarks

The purpose of this paper is actually a promotion of Liu's uncertainty theory [6, 7]. The angle is slightly different from those involving mathematic treatments of high level. In this paper, we first review the well-known "6 sigma rules" of Gaussian distribution in order to trigger a simple question: does "6 sigma rules" have similar form if we change the working environment from Kolmogorov's [1] probability theory into Liu's uncertainty theory [6, 7]? To address such a element but simple question, we review a few simple results of Liu's [7] axiomatic uncertain measure theory, then carry on an investigation of the variance of an uncertain normal
distribution. It is a big surprise that the concept of the uncertain 6 sigma rules and the applications are very delicate. Many deviations in the interval-valued variance or standard deviation are direct consequences of Liu's interval limits [7]. But the interval-valued variance for the uncertain normal distribution revealed that the normal uncertain distribution enjoys a special interval-valued variance, \([\sigma^2_{\xi,u}/2, \sigma^2_{\xi,u}]\), whose lower limit is half of the upper limit, \(\sigma^2_{\xi,u}\).

As the first consequence of the interval-valued variance, then the out control (beyond 6 sigma) is increasing up to two times level than that of the Gaussian regime if picking the upper limit \(\sigma^2_{\xi,u}\) as the variance parameter. Otherwise, the risk level would be much higher up to five times than that of Gaussian 0.27% level.

The curious problem is why should we consider the switching working environment issue? There are many possible answers. For example, in complex system modeling, the uncertainty is hardly to be explained by probability theory. Rather, Liu's uncertainty theory [6, 7] may pave the way to explore the uncertainty in complex system modeling. Finally, we should stress that to apply the uncertain 6 sigma rules depends upon the observational knowledge of the variance or standard deviation of an uncertain distribution if the uncertain normal environment is identified. The right choice is to search the upper bound of the variance sequence if we are able to obtain the sequence.

Acknowledgements

This research is partially supported financially by the South African National Research Foundation (IFR2009090800013 and IFR2011040400096).

References


Danni Guo is a Specialist Scientist at the South African National Biodiversity Institute, Cape Town, South Africa. Her research focus is on spatial modeling under small sample uncertainty as well as fuzzy uncertainty and GIS modeling in biodiversity.

Renkuan Guo is a Emeritus Professor in the Department of Statistical Sciences, University of Cape Town, South Africa. His research interest includes: DEAR (differential equation associated regression) theory; uncertain regression theory with applications in quality and reliability, spatial data modeling and financial market.

Danni Guo is a Specialist Scientist at the South African National Biodiversity Institute, Cape Town, South Africa. Her research focus is on spatial modeling under small sample uncertainty as well as fuzzy uncertainty and GIS modeling in biodiversity.

Renkuan Guo is a Emeritus Professor in the Department of Statistical Sciences, University of Cape Town, South Africa. His research interest includes: DEAR (differential equation associated regression) theory; uncertain regression theory with applications in quality and reliability, spatial data modeling and financial market.