

IMPRECISE RANDOMNESS: AN APPLICATION OF THE MATHEMATICS OF PARTIAL PRESENCE

DHRUBA DAS, dhruadas16@gmail.com

HEMANTA K. BARUAH, hemanta_bh@yahoo.com

Department of Statistics, Gauhati University, Guwahati- 781014, Assam, India.

Abstract: In this article we are going to discuss imprecise randomness using the mathematics of partial presence. The mathematical explanations of imprecise randomness would actually be complete only if it is explained with reference to the Randomness- Impreciseness Consistency Principle. In this article, we have described imprecise randomness with reference to a numerical example of the two sample t- test.

Keywords: *The randomness- impreciseness consistency principle, two sample t- test.*

1. Introduction

If the realizations of a random variable are imprecise in the sense that two independent laws of randomness can define the presence level of values of the variable in a given interval, we would have to deal with the matters using the idea of imprecise randomness (Baruah (2012)). An apparently similar theory, the theory of fuzzy sets, is already in existence since 1965. In the theory of fuzzy sets, it had been accepted that the fuzzy sets do not conform to the classical measure theoretic formalisms. Secondly, it had been agreed upon that given a fuzzy set neither its intersection with its complement is the null set, nor its union with the complement is the universal set. In the Zadehian definition of complementation, fuzzy membership function and fuzzy membership value have been taken to be the same, and that is where the problem lies. Indeed fuzzy membership function and fuzzy membership value are two different things for the complement of a normal fuzzy set (Baruah (1999, 2011)). Instead of saying that the theory of fuzzy sets has been incorrectly explained, Baruah(Baruah (2012))has started the whole process anew naming his finding as

the theory of imprecise sets. Fuzzy randomness in terms of uncertain probabilities has been studied by Buckley (Buckley (2003)) and Buckley and Eslami (Buckley &Eslami (2004))among others. Our approach is however different from their's in the sense that we would be defining imprecise randomness using the Randomness- Impreciseness Consistency Principle together with our definition of complement of an imprecise set.

Baruah has already established(Baruah (2012)) that every law of impreciseness can actually be expressed in terms of two laws of randomness, with randomness defined in the measure theoretic sense. In this article, we are going to discuss about testing statistical hypotheses with reference to imprecise randomness.

In what follows, we shall first discuss about Baruah's Randomness- Impreciseness Consistency Principle andthe complement of an imprecise set. Thereafter we shall discuss the matters with reference to testing an imprecise hypothesis with reference to thetwo sample t- test.

2. The Randomness- Impreciseness Consistency Principle

A normal imprecise number $N = [\alpha, \beta, \gamma]$ is associated with a presence level indicator function $\mu_N(x)$, where

$$\mu_N(x) = \begin{cases} \psi_1(x), & \text{if } \alpha \leq x \leq \beta \\ \psi_2(x), & \text{if } \beta \leq x \leq \gamma \\ 0, & \text{otherwise,} \end{cases}$$

with a constant reference function 0 in the entire real line (Baruah (2012)). Here $\psi_1(x)$ is continuous and non-decreasing in the interval $[\alpha, \beta]$, and $\psi_2(x)$ is continuous and non-increasing in the interval $[\beta, \gamma]$, with

$$\begin{aligned} \psi_1(\alpha) &= \psi_2(\gamma) = 0, \\ \psi_1(\beta) &= \psi_2(\beta) = 1. \end{aligned}$$

The imprecise number would be characterized by $\{x, \mu_N(x), 0 : x \in R\}$, R being the real line. In the Dubois-Prade nomenclature, for a fuzzy number with fuzzy membership function $\mu_N(x)$, $\psi_1(x)$ is called the Left Reference Function, and $\psi_2(x)$ is called the Right Reference Function of the normal fuzzy number. Defining the operation called superimposition of sets and using the Glivenko-Cantelli Theorem (Loeve (1977)) on Order Statistic, Baruah (Baruah (2012)) has established the following result.

Theorem 1: For a normal imprecise number $N = [\alpha, \beta, \gamma]$ with presence level indicator function

$$\mu_N(x) = \begin{cases} \psi_1(x), & \text{if } \alpha \leq x \leq \beta \\ \psi_2(x), & \text{if } \beta \leq x \leq \gamma \\ 0, & \text{otherwise,} \end{cases}$$

such that

$$\begin{aligned} \psi_1(\alpha) &= \psi_2(\gamma) = 0, \\ \psi_1(\beta) &= \psi_2(\beta) = 1. \end{aligned}$$

with constant reference function equal to 0, $\psi_1(x)$ is the distribution function of a random variable defined in the interval

$[\alpha, \beta]$, and $\psi_2(x)$ is the complementary distribution function in the interval $[\beta, \gamma]$. Here, the term random variable has been used in the broader measure theoretic sense. It should be noted here that the notion of probability does not enter in to the measure theoretic definition of a random variable (Rohatgi & Saleh (2001)).

3. The Complement of an Imprecise Number

If a normal imprecise number $N = [\alpha, \beta, \gamma]$ is defined with a presence level indicator function $\mu_N(x)$, where

$$\mu_N(x) = \begin{cases} \psi_1(x), & \text{if } \alpha \leq x \leq \beta \\ \psi_2(x), & \text{if } \beta \leq x \leq \gamma \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} \psi_1(\alpha) &= \psi_2(\gamma) = 0, \\ \psi_1(\beta) &= \psi_2(\beta) = 1. \end{aligned}$$

the complement N^c will have the membership function $\mu_N^c(x)$, where

$$\mu_N^c(x) = 1, \quad -\infty < x < \infty,$$

with the condition that $\mu_N^c(x)$ is to be counted from $\psi_1(x)$, if $\alpha \leq x \leq \beta$, from $\psi_2(x)$, if $\beta \leq x \leq \gamma$, and from 0, otherwise. The definition of the complement of an imprecise set is based on the following axiom:

Axiom 1: For a normal imprecise number $\{x, \mu_N(x), 0 : x \in R\}$ as defined above, the complement $N^c = \{x, \mu_N(x), 1 : x \in R\}$ will have constant presence level indicator function equal to 1, the reference function being $\mu_N(x)$ for $-\infty < x < \infty$.

4. Imprecise Randomness

If the two laws of randomness defining impreciseness are indeed laws of probability, two possibilities can actually be there. When a non-rejectable hypothesis is made imprecise, there may still be a probability that the imprecise hypothesis would actually be found rejectable, the probability of rejection decided by the right reference function. In the same way, if a rejectable hypothesis is made imprecise, there may still be a probability that the imprecise hypothesis would be found non-rejectable, the probability of non-rejection being decided by the left reference function this time (Baruah (2011)).

Assume that X is a random variable following the normal probability law with mean μ and variance unity. Now if the parameter μ is imprecise, with membership defined in $[\mu - \delta, \mu, \mu + \delta]$, we would actually define an infinite number of normal probability density functions with location parameter ranging from $(\mu - \delta)$ to $(\mu + \delta)$ with maximum membership assigned at the value μ . This is where the current definition of imprecise randomness ends.

Assume that we have a sample of n observations x_1, x_2, \dots, x_n from a normally distributed population with mean μ and variance σ^2 . We can then proceed to infer about the population, based on the sample data. Assume further that we have imprecise data and we need to proceed for statistical analysis with reference to imprecise randomness.

The imprecise data are in terms of imprecise numbers around $x_i, i=1,2,\dots,n$ defined as, say,

$$X_i = [x_i - \delta, x_i, x_i + \delta], \delta \geq 0.$$

The analysis can now proceed towards making an imprecise statistical analysis. Without any loss of generality, and for

computational simplicity, such imprecise numbers are usually taken as triangular.

It can be seen that from the distribution function $\psi_1(x)$, for $x_i - \delta \leq x \leq x_i$, we shall get the density function

$$\frac{d}{dx} \psi_1(x) = \varphi_1(x), \text{ say}$$

where

$$\int_{\alpha}^{\beta} \varphi_1(x) dx = 1.$$

In the same way, from the complementary distribution function $\psi_2(x)$, for $x_i \leq x \leq x_i + \delta$ we shall get the density function

$$\frac{d}{dx} (1 - \psi_2(x)) = \varphi_2(x), \text{ say,}$$

where

$$\int_{\beta}^{\gamma} \varphi_2(x) dx = 1.$$

Now, according to Baruah's Randomness – Impreciseness Consistency Principle, a triangular imprecise number of the type

$$X_i = [x_i - \delta, x_i, x_i + \delta]$$

with membership function

$$\mu_{x_i}(x) = \begin{cases} \frac{x - x_i + \delta}{\delta}, & \text{if } x_i - \delta \leq x \leq x_i, \\ \frac{x_i + \delta - x}{\delta}, & \text{if } x_i \leq x \leq x_i + \delta, \\ 0, & \text{otherwise} \end{cases}$$

is in fact defined by two laws of randomness with distribution functions

$$F_1(x) = \frac{x - x_i + \delta}{\delta}, \text{ if } x_i - \delta \leq x \leq x_i$$

and

$$F_2(x) = 1 - \frac{x_i + \delta - x}{\delta}, \text{ if } x_i \leq x \leq x_i + \delta,$$

so that their densities

$$\frac{d}{dx} F_1(x) = \frac{1}{\delta}, \text{ if } x_i - \delta \leq x \leq x_i,$$

and

$$\frac{d}{dx} F_2(x) = \frac{1}{\delta}, \text{ if } x_i \leq x \leq x_i + \delta$$

are uniform.

Accordingly, for imprecise randomness first, there should be a variable following some law of randomness. Secondly, in an interval around every realization of the random variable, there should be impreciseness. If it is presumed that the two laws of randomness are in fact two laws of probability, then the conclusions can be made probabilistically.

5. Two samplet- Test with Imprecise Data

We now cite a numerical example. The gain in weights (in kgs.) of pigs fed on two diets *A* and *B* are given below:

Diet *A*: 49, 53, 51, 52, 47, 50, 52, 53

Diet *B*: 51, 54, 51, 52, 49, 53, 53, 52

Assume that the random samples have been collected from normal populations and the population variances are equal and unknown. Now we want to test whether the two diets differ significantly as regards their effect on increase in weight, i.e., $H_0 : \mu_A = \mu_B$, against the alternative hypothesis, $H_1 : \mu_A \neq \mu_B$.

Under H_0 , the test statistic is given by

$$t = \frac{|\bar{x} - \bar{y}|}{S \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

which follows the Student's- *t* probability law with $(n_1 + n_2 - 2)$ degrees of freedom, where \bar{x} and \bar{y} is the sample means for diet *A* and diet *B* respectively and S^2 is the sample variance.

In our case, the calculated value of *t* is 1.083 which is less than the tabulated value of *t*.e. 2.15 at 5% probability level of significance for 14 degrees of freedom. Therefore we may

conclude that there is no reason to reject the null hypothesis that the diets *A* and *B* does not differ significantly as regards their effect on increase in weight.

Let us now assume that the data are imprecise of the interval type $[x_i - 1, x_i, x_i + 1]$ and that the data are triangular. The random variable *X* of which *x* is a realization in the sample was assumed to be normally distributed. In other words, there is one law of randomness in $[x_i - 1, x_i]$ while there is another law of randomness in $[x_i, x_i + 1]$, both of them being uniform, for a normally distributed realization *x* with mean μ and error variance σ^2 , say. It should be noted here that for probabilistic conclusions based on imprecise random data, we would need to define that these two laws of randomness are indeed two probability laws in the statistical sense.

Thus the gain in weights (in kgs.) of pigs fed on two diets *A* and *B* with triangular membership functions are as follows:

Diet *A*: [48, 49, 50], [52, 53, 54], [50, 51, 52], [51, 52, 53], [46, 47, 48], [49, 50, 51], [51, 52, 53], [52, 53, 54]

Diet *B*: [50, 51, 52], [53, 54, 55], [50, 51, 52], [51, 52, 53], [48, 49, 50], [52, 53, 54], [52, 53, 54], [51, 52, 53]

The null hypothesis would be

$$H_0 : \mu_A = \mu_B$$

where μ_A and μ_B are the imprecise population means concerned.

The alternative hypothesis H_1 would be defined as discussed earlier.

Under H_0 , we finally arrived at an imprecise value of Student's- t statistic for 14 degrees of freedom, with the following presence level indicator function:

$$\mu_i(x) = \begin{cases} \frac{-(-218x^2 + 224) - \sqrt{(-218x^2 + 224)^2 - 4(64x^2 - 224)(201.75x^2 - 56)}}{(128x^2 - 448)}; & 0 \leq x \leq 1.083 \\ \frac{-(100.5x^2 + 672) + \sqrt{(100.5x^2 + 672)^2 - 4(24x^2 + 224)(504 + 28.75x^2)}}{-(48x^2 + 448)}; & 1.083 \leq x \leq \infty \\ 0, & \text{otherwise} \end{cases}$$

Now, according to the Randomness- Impreciseness Consistency Principle this imprecise number gives the following two distribution functions

$$\psi_1(x) = \frac{-(-218x^2 + 224) - \sqrt{(-218x^2 + 224)^2 - 4(64x^2 - 224)(201.75x^2 - 56)}}{(128x^2 - 448)}; \quad 0 \leq x \leq 1.083$$

and

$$(1 - \psi_2(x)) = 1 - \frac{\sqrt{(100.5x^2 + 672)^2 - 4(24x^2 + 224)(504 + 28.75x^2)} - (100.5x^2 + 672)}{-(48x^2 + 448)}; \quad 1.083 \leq x \leq \infty$$

This means, the imprecise number $t = [0, 1.083, \infty]$ with left and right reference functions $\psi_1(x)$ and $\psi_2(x)$ defined in $0 \leq x \leq 1.083$ and $1.083 \leq x \leq \infty$ respectively, would be defined by the two densities $\frac{d}{dx}\psi_1(x)$ and $\frac{d}{dx}(1 - \psi_2(x))$ in the respective ranges. It should be noted here that we have considered the left reference function of the variance at $\alpha = 0.3088$ so that the range of the variance should be positive. Again in computing the value of t , we have also considered the left reference function at $\alpha = 0.5$ for keeping the range of the t as 0 to ∞ .

In the crisp or non- imprecise situation of the above example, we would have concluded that there is no reason to

reject the null hypothesis of equality of the mean weights at 5% probability level of significance as the data dependent value of $t (= 1.083)$ is smaller than the theoretical value of $t (= 2.15)$ for 14 degrees of freedom. We now proceed to look in to the matters of making an imprecise conclusion statistically. The tabulated non- imprecise value of t at 5% level of significance for 14 degrees of freedom is 2.15, which lies between 1.083 to ∞ and the right of this tabulated value of t , i.e. 2.15, the area under the probability function of t is 0.025. 2.15 is on that part of the interval on which the right reference function is defined. The membership for $t = 2.15$ is $\psi_2(2.15)$. Now, in our perspective the probability density function concerned would be given by

$$\frac{d}{dx}(1-\psi_2(x)) = -\frac{25536x}{(48x^2 + 448)^2} + \frac{(3652320x^3 + 27295744)}{(48x^2 + 448)^2 \sqrt{7340.25x^4 + 60928x^2}}; \quad 1.083 \leq x \leq \infty$$

Therefore, the probability that $t \geq 2.15$ would be the area under this probability density function for $t \geq 2.15$, which is the area of the right tail beyond 2.15. The area of the left tail from 1.083 to 2.15 is $(1 - \psi_2(2.15))$. Thus the area of the right tail is $\psi_2(2.15)$, which is no thing but the membership value of t at 2.15. Therefore $\psi_2(2.15) = 0.708158$ is the probability that the imprecise null hypothesis would have to be rejected at 5% probability level of significance. In other words, when a non rejectable hypothesis is made imprecise, there may still be a probability that the imprecise hypothesis would actually be found rejectable. In the same way, if a rejectable hypothesis is made imprecise, there may still be a probability that the imprecise hypothesis would be found non-rejectable, the probability of non rejection being decided by the left reference function this time.

6. Conclusions:

Two laws of randomness are necessary and sufficient to define the partial presence of an element in a normal imprecise number. In this article, based on the Randomness-Impreciseness Consistency Principle, we have forwarded the definition of imprecise randomness. In testing of imprecise hypothesis, we deal with the alternative hypothesis which is the complement of the imprecise null hypothesis. We have shown that when a non rejectable hypothesis is made imprecise, there may still be a probability that the imprecise hypothesis would actually be found rejectable. In the same way if a rejectable hypothesis is made imprecise, there may still be a probability

that the imprecise hypothesis would be found non-rejectable.

References:

1. Baruah, H. K. (1999). Set Superimposition and Its Application to the Theory of Fuzzy Sets. *Journal of the Assam Science Society*, 40(1 & 2), 25 – 31.
2. Baruah, H. K. (2011). The theory of fuzzy sets: beliefs and realities. *International Journal of Energy Information and Communications*, 2(2), 1-22.
3. Baruah, H. K. (2012). An Introduction to the Theory of Imprecise Sets: The Mathematics of Partial Presence. *Journal of Mathematical and Computational Sciences*, 2(2), 110 – 124.
4. Buckley, J. J. (2003). Uncertain probabilities III: the continuous case. *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, 8(3), 200 – 206.
5. Buckley, J. J., & Eslami, E. (2004). Uncertain probabilities II: the continuous case. *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, 8(3), 193 – 199.
6. Loeve, M. (1977). *Probability Theory*. New York: Springer Verlag.
7. Rohatgi, V. K., & Saleh, A. K. (2001). *An Introduction to Probability and Statistics*. Singapore: Second Edition, Wiley Series in Probability and Statistics, John Wiley and Sons (Asia) Pte Ltd.