Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.13, 2013
www.iiste.org

A Method for Constructing Fuzzy Test Statistics with Application

Assel S. Mohammad
Al – Kindy Medical Collage, University of Baghdad,Iraq
nbrhms@yahoo.com

Abstract:

The scale parameter (θ) play an important role in statistical inference and in statistical testing of hypotheses ,but sometimes when the available observation about this parameter ,especially when (θ) represent the percentage of defective in production ,or mean time to failure of certain product ,imprecise information about parameter of the distribution under hypotheses testing lead to test the hypothesis with fuzzy concepts .this paper deals with problem of testing hypothesis when the hypotheses are fuzzy and data are crisp. We first introduce the approach to testing fuzzy hypothesis ,when the parameter θ represent percentage of defective in production and it is considered random variable having prior distribution $g(\theta)$,and each inspected item is considered either good or defective so the distribution of (n) inspected units is binomial with parameters (n,θ) .The data used are taken from certain industrial company in Baghdad ,Iraq. The fuzzy hypothesis test here were done by two methods, the first is the fuzzy Bayesian sampling test and the second depend on the procedure of using the $[\alpha - cut]$ of confidence interval of the parameter θ , $\tilde{X}(\alpha)$ .first the introduction is given and second some preliminaries about fuzzy ,then we explain fuzzy hypothesis and finally the application.

Keywords: Testing fuzzy hypothesis, Membership function, Prior distribution,Posterior distribution.

1. Introduction

The statistical analysis in its normal form depend on crispness of data, random variables, hypothesis, decision rules and parameters estimations and testing.

The Bayesian approach is one of the important methods which is used in statistical analysis in estimation and testing hypothesis. Thomas Bayes in 1967 discussed the estimation of the parameter θ which represent the percentage of defective in products and considered to be random variable varied from lot to lot. Casals &Gil in 1986 studied the problem of testing statistical hypothesis with fuzzy sets and vague data. Also Casals and Gill in 1994 introduce Bayesian sequential test for fuzzy parametric problem.

Arnold in 1996 gives an approach of fuzzy hypothesis testing .In 2001 Holenam gives a fuzzy logic generalization of a data mining.Viertl in 2006 introduce some methods to construct confidence intervals and statistical tests for fuzzy data. Many papers deals with problems of testing fuzzy hypothesis using fuzzy data and work on constructing a test statistic like Arefi&Taheri 2011and Bukley 2004 and Denoeus& Masson 2005.Our research deals with introducing a method for testing fuzzy hypothesis about the percentage of defective in production process and how to take a decision for acceptance or rejection of $H_0(\theta)$ and $H_1(\theta)$ depending on a statistical test for fuzzy Bayesian statistical hypothesis which to be solved using numerical integral method.

2. Fuzzy hypothesis

Any hypothesis about the parameter θ ,written in the form $(\tilde{H}:\theta)$,H is called a fuzzy hypothesis ,where $H:\theta \to [0,1]$ is a fuzzy subset of parameter space Θ ,with membership function H .The ordinary hypothesis is $(H_0:\theta \in \theta_0)$ is called a fuzzyhypothesis with the membership function $(H_0 : I_{\theta_0})$ .here are some required definitions.

Definition1:

The fuzzy hypothesis $(\tilde{H}:\theta \; is \; H)$ be such that

1- H is a monotone function of θ
2- There exists $\theta_1 \in \Theta$ such that $H(\theta)=1$ for $\theta \geq \theta_1$ or ($\theta \leq \theta_1$)

3- The range of H contains the interval $(0,1]$, then $\tilde{H}$ is called a one sided fuzzy hypothesis. If fuzzy hypothesis $(\tilde{H}:\theta \; is \; H)$ be such that there exist an interval $[\theta_1,\theta_2] \subset \Theta$ such that $H(\theta)=1$ for $\theta \in [\theta_1,\theta_2]$ and $\inf[\theta:\theta \in \Theta] < \theta_1 \leq \theta_2 < \sup[\theta:\theta \in \Theta]$. And H is an increasing function for

$\theta \le \theta_1$ and is decreeing for $\theta \ge \theta_1$ and the range of H contains the interval $(0,1]$,then $\widetilde{H}$ is called two – sided fuzzy hypothesis.

Definition2:

The ordinary hypothesis $(H_0 : \theta \in \theta_0)$ is a fuzzy hypothesis with membershipfunction $(H_0 = I_{\theta_0})$.

In testing simple or composed hypothesis about parameter and there is indicator define the range parameter space we can use NeymanPearson or maximum likelihood ratio test to obtain the test statistics based on sample information to take a decision for accept or reject. But in case of fuzzy hypothesis the hypothesis about θ is $H : \theta\ H(\theta)$ a membership function on θ ,i.e. it is a function from θ to $[0,1]$.

Let $X = (X_1, X_2, ..., X_n)$ be a random sample that observed values $x = (x_1, x_2, ..., x_n)$ and X_i have probability function $f(x_i|\theta)\theta \in \Theta,$ is unknown parameter and have prior distribution $\pi(\theta)$.

Let $H_0(\theta), H_1(\theta)$ be two membership functions, and the range of H contains the interval $(0,1]$.It is requires to test;

$H_0 : \theta\ is\ H_0(\theta)$

$H_1 : \theta\ is\ H_1(\theta)$

Using Bayesian test we need the Loss function $L(\theta, a), a \in H$ is the set of all possible decision and $L(\theta, a) : \theta \times A \to R$ is a loss function from taking decision a according to state of nature (θ).

Let D be the set of all decision functions which define ($R^n\ on\ A$), then the risk function due to the wrong decision is defined to be the expected Loss:

$$R(\theta, d) = E\big(L(\theta, d(X))\big) \quad (1)$$

Is the risk function from taken d about θ.

When θ is random variable, have $\pi(\theta)$, then the posterior distribution of θ given the sample $x = (x_1, x_2, ..., x_n)$ is;

$$\pi(\theta|x) \propto \pi(\theta) f(x|\theta)$$

And $f(x,\theta) = \pi(\theta) f(x|\theta)$

$= f_X(x)\pi(\theta|x)$

And Bayes Risk due to (d)

$$R(\pi, d) = E\big(R(\theta, d)\big) \quad (2)$$

Here we need to define the Bayes test according to;

1- Bayes test without Loss function

Here we want to test the fuzzy hypothesis $H_0(\theta)$ against $H_1(\theta)$,according to sample information $f(x|\theta)$ and $\pi(\theta)$,the rule here depend on using membership function instead of indicator function.

Let $\theta = \theta_0 \cup \theta_1$ be the space of parameter,and

$$H_0(\theta) = \begin{cases} 1 & if\ \theta \in \theta_0 \\ 0 & if\ \theta \in \theta_1 \end{cases} \quad (3)$$

$$H_1(\theta) = \begin{cases} 0 & if\ \theta \in \theta_0 \\ 1 & if\ \theta \in \theta_1 \end{cases} \quad (4)$$

And

$$\int \pi(\theta|x) H_0(\theta) d\theta = \int_{\theta_0} \pi(\theta|x) d\theta$$

$$= pr(\theta \in \theta_0|x)$$

$$\int \pi(\theta|x)H_1(\theta)d\theta = \int_{\theta_1} \pi(\theta|x)d\theta$$
$$= pr(\theta \in \theta_1|x)$$

H_0 is accepted when $pr(\theta \in \theta_0|x) \geq pr(\theta \in \theta_0|x)$

Some researcher introduce a factor called degree of certainty about our decision defined as $D = \frac{\alpha_0}{\alpha_0 + \alpha_1}$

Where;

$$\alpha_0 = \int \pi(\theta|x)H_0(\theta)d\theta$$
$$\alpha_1 = \int \pi(\theta|x)H_1(\theta)d\theta$$

ii- Bayes test of fuzzy hypothesis using loss function

The test depend on loss function which defined on membership function $H_0(\theta)$, $H_1(\theta)$,i.e:

$L(\theta, a_0) = a(\theta)[1 - H_0(\theta)]$ (5)

$L(\theta, a_1) = b(\theta)[1 - H_1(\theta)]$ (6)

$a(\theta), b(\theta)$ ordinary positive function defined on θ and choosing of it depend on the sensitivity of wrong decision for rejection or acceptance.

For this test H_0 is accepted if

$$\int a(\theta)\,[1 - H_0(\theta)]\pi(\theta|x)d\theta \leq \int b(\theta)\,[1 - H_1(\theta)]\pi(\theta|x)d\theta$$

Let $a(\theta) = c_{11}$, $b(\theta) = c_1$ be constants.

Then the fuzzy Bayes test H_0 is accepted if

$$\frac{1 - \int H_1(\theta)\pi(\theta|x)d\theta}{1 - \int H_1(\theta)\pi(\theta|x)d\theta} \geq \frac{c_{11}}{c_1} \qquad (7)$$

c_1 may represent the cost or the loss when type I error is happen (rejecting true hypothesis)

Table (1): percentage of defectives

and c_{11} is the loss or cost due to type II error (which is the probability of accepting false hypothesis).

The distribution of percentage of defective of 120 Lots ,each of size production (2000) ,from certain product of Iraqi –industrial are;

$p_i\%$	0.1	0.2	0.3	0.4	0.5	0.6	0.7
f_i	6	20	34	29	18	9	4

Which found to be Beta with $a = 3, b = 4$ with estimated $\overline{p} = 0.4285$

The $p.d.f$ of p_i which is represented by $g(\theta)$ or $\pi(\theta)$

$$\pi(\theta) = \begin{cases} 60\theta^2(1-\theta)^3 & 0 \leq \theta \leq 1 \\ 0 & o.w \end{cases}$$

And we find the estimated average of percentage of defective in this industrial state is $(\overline{p} = 0.0425)$,this percentage does not consist with (AOQL:Average outgoing quantity level,equal 2%)and the produced units are important,so the company work on applying the testing of fuzzy hypothesis were percentage of defective is about $(0.2 \leq \theta \leq 0.43)\%$.

Therefore the fuzzy hypothesis is;

$H_0 : (0.2 \leq \theta \leq 0.4)\%$

$H_1 : \theta \; is \; away \; from \; 0.4$

Now we shall find the posterior distribution of $h(\theta|y)$,since each produced unit inspected and it is either good or defective so the distribution of number of defective is Bernoulli and the $(y=\sum_{i=1}^{n} x_i)$ (sum of defectives in the sample n) is random variable follow binomial distributing (n,θ) .

Therefore;

$$h(\theta|y)=\frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta}$$

$$=\frac{C_y^n\theta^y(1-\theta)^{n-y}60\theta^2(1-\theta)^3}{\int_0^1 60C_y^n\theta^{y+2}(1-\theta)^{n+3-y}d\theta}$$

$$=\frac{\theta^{y+2}(1-\theta)^{n+3-y}}{Beta(y+3,n+4-y)}$$

This indicates that;

$$h(\theta|y)\sim Beta\,(y+3,n+4-y)$$

Now the membership function under H_0 and under H_1 is;

$$H_0(\theta)=\begin{cases}5\theta-1 & 0.2\le\theta\le 0.4\\ 3-5\theta & 0.4\le\theta<0.6\end{cases}$$

$$H_1(\theta)=\begin{cases}1-2.5\,\theta & 0\le\theta\le 0.4\\ \frac{5\,\theta}{3}-\frac{2}{3} & 0.4\le\theta<1\end{cases}$$

According to fuzzy hypothesis and membership function, H_0 is rejected if;

$$\int H_0\,(\theta)\pi(\theta|y)d\theta\le\int H_1\,(\theta)\pi(\theta|y)d\theta$$

$\forall\theta\in H_0(\theta)$ $\qquad\forall\theta\in H_1(\theta)$

$I_1 I_2$

After solving I_1 under $H_0(\theta)$ and I_2 under $H_1(\theta)$

If $I_1\le I_2$ then reject $H_0(\theta)$ otherwise accept $H_0(\theta)$

$$I_1=\int H_0\,(\theta)\pi(\theta|y)d\theta$$

$\forall\theta\in H_0(\theta)$

$$=\int_{0.2}^{0.4}\frac{(5\theta-1)\theta^{y+2}(1-\theta)^{n+3-y}}{Beta(y+3,n+4-y)}d\theta+\int_{0.4}^{0.6}\frac{(3-5\theta)\theta^{y+2}(1-\theta)^{n+3-y}}{Beta(y+3,n+4-y)}d\theta$$

And

$$I_2=\int H_1\,(\theta)\pi(\theta|y)d\theta$$

$\forall\theta\in H_1(\theta)$

$$=\int_{0}^{0.4}\frac{(1-2.5\theta)\theta^{y+2}(1-\theta)^{n+3-y}}{Beta(y+3,n+4-y)}d\theta+\int_{0.4}^{1}\frac{1/3(5\theta-2)\theta^{y+2}(1-\theta)^{n+3-y}}{Beta(y+3,n+4-y)}d\theta$$

Now for I_1

$$I_1=\frac{1}{k}[\int_{0.2}^{0.4}5\,\theta^{y+3}\,(1-\theta)^{n+3-y}d\theta-\int_{0.2}^{0.4}\theta^{y+2}\,(1-\theta)^{n+3-y}d\theta$$

$$+\int_{0.4}^{0.6} 3\,\theta^{y+2}(1-\theta)^{n+3-y}\,d\theta - 5\int_{0.2}^{0.4} \theta^{y+3}(1-\theta)^{n+3-y}\,d\theta]$$

for $n = 5,10$, I_1 can be solved numerically

Also for I_2

$$I_2 = \frac{1}{k}[\int_0^{0.4}\theta^{y+2}(1-\theta)^{n+3-y}d\theta - 2.5\int_0^{0.4}\theta^{y+3}(1-\theta)^{n+3-y}d\theta + \int_{0.4}^{1}5/3\theta^{y+3}(1-\theta)^{n+3-y}d\theta - \int_{0.4}^{1}2/3\theta^{y+2}(1-\theta)^{n+3-y}d\theta]$$

$$k = Beta(y+3, n+4-y)$$

Also I_2 can be solved numerically using trapezoidal method or Simpsons.

Then the results for I_1, I_2 are computed ,when $I_1 \le I_2$,then the Bayesdecision is work on rejecting H_0 and accepting H_1 otherwise is true.

Table (2):results of integral $I_1 \& I_2$ $when\ n = 5, n = 10$

n	5		
y	I_1	I_2	**decision**
0	**0.2894**	**0.4115**	**Reject H_0**
1	**0.4391**	**0.2814**	**Accept**
2	**0.4843**	**0.2253**	**Accept**
3	**0.4044**	**0.2479**	**Accept**
4	**0.2575**	**0.3318**	**Reject H_0**
5	**0.1224**	**0.4508**	**Reject H_0**
n	**10**		
y	I_1	I_2	**decision**
0	**0.1239**	**0.5624**	**Reject H_0**
1	**0.25816**	**0.42672**	**Reject H_0**
2	**0.40779**	**0.31054**	**Accept**
3	**0.51868**	**0.22817**	**Accept**
4	**0.54853**	**0.19087**	**Accept**
5	**0.48969**	**0.2008**	**Accept**
6	**0.37035**	**0.25008**	**Accept**
7	**0.23587**	**0.3256**	**Reject H_0**
8	**0.12461**	**0.41511**	**Reject H0**
9	**0.053215**	**0.51055**	**Reject H0**
10	**0.017639**	**0.60797**	**Reject H0**

Conclusion

In practical problems, we may face fuzzy information about observation and about parameter, so we may face fuzzy hypothesis rather than crisp hypothesis, for example, when we are interested in evaluating average of percentage of defectives for testing product produced by a factory, and the information are uncertain, this lead to perform the fuzzy test rather than crisp. The present work introduce a fuzzy test for the hypothesis about the percentage of defective in product of (120) lot of some product produced by some Iraqi industries. The hypothesis tested depend on membership function $H_0(\theta), H_1(\theta)$, where the underling distribution is Beta – Binomial. After formulating the test statistics, which depend on the fuzzy Bayes test H_0 against H_1. The values of integral I_1 and I_2 are computed numerically using Trapezoidal method, the results are explained for $n = 5, n = 10$, when $I_1 \le I_2$, the Bayes decision work on rejecting H_0 and accepting H_1, the results are explained in table (2). Also it is necessary to remark that the membership function, here under $H_0(\theta)$, are constructed depend on the mean value of estimated average of defective ($\overline{p} = 0.4285$), which is indicated by θ for $H_0(\theta)$ and $H_1(\theta)$.

References

[1] Arefi, M., and Taheri, S. M. (2011). Testing fuzzy hypotheses using fuzzy data basedfuzzy test statistic. Journal of Uncertain Systems, 5, 45-61.

[2] Arnold B.F (1996),"An approach to fuzzy hypothesis testing",metrika,4,4:119-126.
[3] Buckley,J.J(2005),"fuzzy statistics hypothesis testing",soft computing:512-518.
[4] Chachi, J., and Taheri, S. M. (2011). Fuzzy confidence intervals for mean of Gaussian fuzzy random variables. Expert Systems with Applications, 38, 5240-5244.
[5] Colubi, A., and Gonz´alez-Rodr´ıguez, G. (2007). Triangular fuzzification of random variables and power of distribution tests: empirical discussion. Computational Statistics and Data Analysis, 51, 4742-4750.
[6] Filzmoser, P., and Viertl, R. (2004). Testing hypotheses with fuzzy data: the fuzzy pvalue.Metrika, 59, 21-29.
[7] Hryniewicz, O. (2006). Possibilistic decisions and fuzzy statistical tests.Fuzzy Sets andSystems, 157.
[8] Jalal Chachi1, Seyed Mahmoud Taheri1 and ReinhardViertl, (2012), " Testing Statistical Hypotheses Based on Fuzzy Confidence Intervals " AUSTRIAN JOURNAL OF STATISTICS Volume 41 (2012), Number 4, 267–286.
[9] Laureano Rodríguez, Gladys Casas, Ricardo Grau, and YailenMartínez, (2007), " Fuzzy Scan Method to Detect Clusters " International Journal of Biological and Life Sciences 3:2 2007.
[10]Parchami, A., Taheri, S. M., and Mashinchi, M. (2010). Testing fuzzy hypotheses based on vague observations: a p-value approach. Statistical Papers, 51, 209-226.
[11]Shang-Ming Zhou, Member, IEEE, and John Q. Gan, (2007), " Constructing L2-SVM-Based Fuzzy Classifiers in High-Dimensional Space With Automatic Model Selection and Fuzzy Rule Ranking " IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 15, NO. 3, JUNE 2007.
[12] Taheri.S.M(2003),"Trends in fuzzy statistics",Aust.J.Stat,32(3):239-257.
[13] Taheri.S.M and Behboodian .J(2005),"A Bayesian approach to fuzzy hypothesis with fuzzy data",Italian Journal of pure and Applied Mathematics,19,139-154.
[14] Viertl, R. (2006). Univariate statistical analysis with fuzzy data. Computational Statistics and Data Analysis, 51, 133-147.
[15]Viertl, R. (2011). Statistical Methods for Fuzzy Data.Chichester: John Wiley and Sons.
[16] Wu,H.Ch,(2005),"Statistical hypothesis testing for fuzzy data",Information Sciences,vol.175,pp.30-56.
[17] Wu, H. C. (2005). Statistical hypotheses testing for fuzzy data. Information Sciences, 175, 30-57.
[18] Wu, H. C. (2009). Statistical confidence intervals for fuzzy data. Expert Systems with
Applications, 36, 2670-26760.
[19] Zadeh,L.A,(1965)"Fuzzy sets "Information and Control,vol.8,pp.338-353.
[20] Zimmermann, H. J. (2001). Fuzzy Set Theory and its Applications (4th ed.). Boston: Kluwer Academic Publishers.