Scientific
Research

Sequential Probability Ratio Test of Correlation Coefficient Using Fuzzy Hypothesis Testing

Sevil Bacanli, Duygu Icen

Department of Statistics, Hacettepe University, Ankara, Turkey
Email: sevil@hacettepe.edu.tr

ABSTRACT

One of the important fields in statistics is testing hypothesis of correlation coefficient. The extension of the idea of testing correlation to fuzzy hypothesis is of great interesting. In this study, we examined the use of fuzzy hypothesis testing approach for the Sequential Probability Ratio Test (SPRT) of correlation coefficient. Use of fuzzy hypothesis testing for correlation coefficient with SPRT is illustrated by an example.

Keywords: Correlation; Sequential Probability Ratio Test; Fuzzy Hypothesis Testing

1. Introduction

Correlation coefficient (ρ) is found to have applications in some area such as medicine, genetics, psychology, numerical taxonomy and quality control [1]. In these areas, the data is obtained sequentially. Thus it is advantageous to use sequential tests.

Wald [2] proposed SPRT which is designed to decide between two simple hypotheses. SPRT require substantially fewer observations than fixed sample size test.

The SPRT for the correlation coefficient has been investigated by Choi [3], Kowalski [1] and Pradhan and Sathe [4]. Choi [3] developed a SPRT which is essentially a test for the Cauchy parameter. Kowalski [1] derived a SPRT under the bivariate normal distribution.

Pradhan and Sathe [4] proposed an unbiased estimator of correlation coefficient. This estimator suggested a SPRT for "ρ" which reduces to an SPRT for the Bernoulli parameter θ, which is a simple function of, but two pairs of observations need each step. Pradhan and Sathe's SPRT is more efficient than Kowalski's and Choi's tests [4].

There are several components for making decisions in traditional statistics such as crispness of data, random variables, hypothesis testing, decision rules and so on. There have been several attempts to analyze above situations with fuzzy set theory developed by Zadeh [5]. One of them is fuzzy hypothesis testing procedure. In fuzzy hypothesis testing the sample data and hypotheses are either crisp or fuzzy. In this study we focus on fuzzy hypothesis with crisp observations.

Several authors have proposed SPRT for fuzzy case. Neyman Pearson lemma for fuzzy hypothesis testing was given by Taheri and Behboodian [6]. Torabi and Behboodian [7] gave the SPRT for fuzzy hypothesis testing with fuzzy observations. Talukdar and Baruah [8] have fuzzified SPRT. Torabi and Mirhosseini [9] introduced a SPRT for fuzzy hypothesis testing. They developed fuzzy SPRT for testing the mean of Normal distribution with known variance and the parameter of Bernoulli distribution (θ) Akbari [10] proposed a new approach for SPRT of fuzzy hypothesis under density probability function. Testing hypothesis concerning correlation coefficient with fuzzy data was investigated by Leu *et al*. [11].

In this study, we apply SPRT to test fuzzy hypothesis for the correlation coefficient of a bivariate normal distribution. Then this method is demonstrated by an example.

This paper is organized as follows: in Section 2, SPRT for the correlation coefficient [4] is given. In Section 3, some of the basic concepts of fuzzy hypothesis testing are defined and reviewed. In Section 4, Torabi and Mirhosseini's SPRT approach for Bernoulli parameter is given [9]. In addition we apply this fuzzy SPRT to test fuzzy hypothesis for the correlation coefficient. Then an example and conclusion are presented in Sections 5 and 6.

2. Sequential Test for the Correlation Coefficient

Let the independent random variable $(X_i, Y_i), i = 1, 2, \cdots$

be from bivariate normal distribution. Than the correlation coefficient is given as

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}} \qquad (1)$$

Let the hypothesis to be tested be

$$\begin{aligned} H_0 &: \rho = \rho_0 \\ H_1 &: \rho = \rho_1 \ (\rho_1 > \rho_0) \end{aligned} \qquad (2)$$

where ρ is a positive or negative constant $(-1 \le \rho \le 1)$. Pradhan and Sathe [4] gave Theorem 1 which establishes a basic relationship between correlation coefficient (ρ) and the Bernoulli parameter (θ)

Theorem 1: *Let the independent random variables* $(X_i, Y_i), i = 1, 2, \cdots$ *be from bivariate Normal distribution* $\left(0, 0, (\sigma_X)^2, (\sigma_Y)^2, \rho\right)$. *Then,*

$$P(X_1 Y_1 + X_2 Y_2 > 0) = (1 + \rho)/2. \qquad (3)$$

In the sequential testing procedure, it is taken two pairs of observations at each step. At the *i*th step it is defined as,

$$U_i = \begin{cases} 1, \text{if } (X_{2i-1} Y_{2i-1} + X_{2i} Y_{2i} > 0) \\ 0, \text{otherwise} \end{cases}, i = 1, 2, \cdots \quad (4)$$

Then U_i follows Bernoulli distribution with

$$P(U_i = 1) = \theta = (1 + \rho)/2. \qquad (5)$$

Hence hypotheses for the correlation coefficient (2) can be written as,

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned} \qquad (6)$$

and SPRT for the Bernoulli parameter can be applied [2].

For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1 (\theta_1 > \theta_0)$ SPRT is defined as follows:

Let $U_1, U_2, \cdots$ be a sequence i.i.d. rv's from Bernouilli distribution Ber (θ) where $0 < \theta < 1$,

$$f(u, \theta) = \begin{cases} \theta, & u = 1 \\ 1 - \theta, u = 0 \end{cases} \qquad (7)$$

then the likelihood ratio is,

$$\begin{aligned} L &= \prod_{i=1}^{n} \frac{f(u_i, \theta_0)}{f(u_i, \theta_1)} = \frac{\theta_0^{u_1 + \cdots + u_n} (1 - \theta_0)^{n - (u_1 + \cdots + u_n)}}{\theta_1^{u_1 + \cdots + u_n} (1 - \theta_1)^{n - (u_1 + \cdots + u_n)}} \\ &= \frac{\theta_0^{\sum_{i=1}^{n} u_i} (1 - \theta_0)^{n - \sum_{i=1}^{n} u_i}}{\theta_1^{\sum_{i=1}^{n} u_i} (1 - \theta_1)^{n - \sum_{i=1}^{n} u_i}} \end{aligned} \qquad (8)$$

By taking logarithm and simplifying, Equation (8) can be written as,

$$\ln L = \sum_{i=1}^{n} z_i \qquad (9)$$

where z_i is taken as following

$$\begin{aligned} z_i &= u_i \ln \frac{\theta_0}{\theta_1} + (1 - u_i) \ln \frac{1 - \theta_0}{1 - \theta_1} \\ z_i &= u_i \left(\ln \frac{\theta_0}{\theta_1} - \ln \frac{1 - \theta_0}{1 - \theta_1} \right) + \ln \frac{(1 - \theta_0)}{(1 - \theta_1)} \end{aligned} \qquad (10)$$

Hence

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} u_i \left(\ln \frac{\theta_0}{\theta_1} - \ln \frac{1 - \theta_0}{1 - \theta_1} \right) + n \ln \frac{(1 - \theta_0)}{(1 - \theta_1)}. \quad (11)$$

After each observation is computed the $\sum_{i=1}^{n} z_i$ and compared with $\ln A = \ln \left[\alpha/(1 - \beta) \right]$ and $\ln B = \ln \left[(1 - \alpha)/\beta \right]$, where α is the type I and β is the type II error probability. Then one of the following decisions is reached

1) If $\sum_{i=1}^{n} z_i \le \ln A$, reject H_0 and stop sampling.

2) If $\ln A < \sum_{i=1}^{n} z_i < \ln B$ continue sampling.

3) If $\sum_{i=1}^{n} z_i \ge \ln B$ accept H_0 and stop sampling.

Let N be the random variable denoting the sample size of the SPRT. The SPRT with α and β, expected sample size when the H_0 is true,

$$E(N/H_0) \approx \frac{(1 - \alpha) \ln B + \alpha \ln A}{E(z; \theta_0)}, \qquad (12)$$

otherwise, when H_1 is true,

$$E(N/H_1) \approx \frac{\beta \ln B + (1 - \beta) \ln A}{E(z; \theta_1)}. \qquad (13)$$

where $E(z, \theta)$ is defined as from Equation (10) $(\theta = (1 + \rho)/2)$

$$E(z, \theta) = \theta \left(\ln \frac{\theta_0}{\theta_1} - \ln \frac{1 - \theta_0}{1 - \theta_1} \right) + \ln \frac{(1 - \theta_0)}{(1 - \theta_1)}. \quad (14)$$

Pradhan-Sathe's [4] SPRT need two pairs of observations each step. Therefore, the expected sample size of this SPRT is given by $E(N; \rho) = 2E(N/H_i) + 2$, for $i = 0, 1$.

3. Fuzzy Hypothesis Testing

We may face a fuzzy hypothesis in practice in many areas [6,9,12]. Hence, in this section we briefly introduce some definitions about fuzzy hypothesis testing (FHT).

Preliminaries

Some preliminaries are given below for testing fuzzy

hypotheses.

Definition 3.1 Any hypothesis of the form "$H : \theta$ is $H(\theta)$" is called fuzzy hypothesis (FH), where "$H : \theta$ is $H(\theta)$" implies that θ is in a fuzzy set of Θ (the parameter space) with membership function $H(\theta)$ i.e. a function from Θ to $[0,1]$.

Note that the ordinary hypothesis $H_i : \theta \in \Theta_i$ is a fuzzy hypothesis with membership function $H(\theta) = 1$ at $\theta \in \Theta_i$, and zero otherwise, i.e., the indicator function of the crisp set $\Theta_i, i = 0,1$ [6,7]

Definition 3.2 In FHT with crisp data, the main problem is testing

$$H_0 : \theta \text{ is } H_0(\theta)$$
$$H_1 : \theta \text{ is } H_1(\theta) \tag{15}$$

according to a random sample $X(X_1, \cdots, X_n)$ from a parametric population with the probability density function (PDF) $f(x;\theta)$. It follows that the PDF of the random sample X is [12]

$$f(X,\theta) = \prod_{i=1}^{n} f(x_i;\theta). \tag{16}$$

Definition 3.3 The pseudo-membership function of $H_j(\theta)$ is defined by

$$H_j^*(\theta) = H_j(\theta) \bigg/ \int_\theta H_j(\theta) \mathrm{d}\theta, j = 0,1. \tag{17}$$

Substitute $\int$ by $\sum$ in discrete cases; i.e. in the case that Θ has just uncountable values [9].

Definition 3.4 Let X be a random variable with PDF $f(x;\theta)$. Under the hypothesis $H_j(\theta), j = 0,1$, the weighted probability density function (WPDF) of X is defined by

$$f_j(x) = \int_\theta H_j^*(\theta) f(x;\theta) \mathrm{d}\theta, \tag{18}$$

i.e., the expected value of $f(x;\theta)$ over $H_j^*(\theta), j = 0,1$. If X is a random sample from PDF $f(\cdot, \theta)$, then the joint WPDF of X is defined by

$$f_j(X) = \prod_{i=1}^{n} f_j(x_i). \tag{19}$$

Then the weighted likelihood ration is defined as [6]

$$L_w = \prod_{i=1}^{n} \frac{f_0(x_i)}{f_1(x_i)}. \tag{20}$$

4. Fuzzy Hypothesis Testing for the Sequential Probability Ratio Test of Correlation Coefficient

It is known that correlation coefficient is related to the Bernoulli parameter $(\theta = (1+\rho)/2)$ [4]. In the sense of this information which is given in Section 1 with Theorem 1, hypothesis testing for correlation coefficient with

Equation (2) can be written as

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1. \tag{21}$$

Therefore, Torabi and Mirhosseini's [9] fuzzy SPRT for the Bernoulli parameter can be applied. For testing (21) fuzzy SPRT is defined as follows:

Let $U_1, U_2, \cdots$ be a sequence of i.i.d. r.v.'s from Bernoulli distribution $(\mathrm{Ber}(\theta))$,

$$f(u, \theta) = \theta^u (1-\theta)^{1-u}, u = 0,1, 0 < \theta < 1. \tag{22}$$

We want to test

$$H_0 : \theta \approx \theta_0 (\theta \text{ is } H_0(\theta))$$
$$H_1 : \theta \approx \theta_1 (\theta \text{ is } H_1(\theta)) \tag{23}$$

with the membership functions [9]

$$H_j(\theta) = \theta^{\alpha_j - 1} (1-\theta)^{\beta_j - 1}, \forall \theta \in (0,1),$$
$$\theta_j = \alpha_j / (\alpha_j + \beta_j), j = 0,1. \tag{24}$$

The pseudo-membership function of $H_j(\theta)$ is defined from Definition 3.3,

$$H_j^*(\theta) = H_j(\theta) \bigg/ \int_\theta H_j(\theta) \mathrm{d}\theta, j = 1,2$$

$$= \theta^{\alpha_j - 1} (1-\theta)^{\beta_j - 1} \bigg/ \int_0^1 \theta^{\alpha_j - 1} (1-\theta)^{\beta_j - 1} \mathrm{d}\theta$$

$$= \theta^{\alpha_j - 1} (1-\theta)^{\beta_j - 1} \bigg/ \frac{\Gamma(\alpha_j)\Gamma(\beta_j)}{\Gamma(\alpha_j + \beta_j)} \tag{25}$$

$$= \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \theta^{\alpha_j - 1} (1-\theta)^{\beta_j - 1},$$

$$j = 0,1; \alpha_j > 0; \beta_j > 0$$

Hence $H_j^*(\theta)$ is Beta distribution. The weighted probability density function of θ is defined by (Definition 3.4)

$$f_j(u) = \int_\theta H_j^*(\theta) f(u;\theta) \mathrm{d}\theta$$

$$= \begin{cases} \beta_j / (\alpha_j + \beta_j), u = 0 \\ \alpha_j / (\alpha_j + \beta_j), u = 1 \end{cases} \tag{26}$$

Hence the FHT (Equations (23)) is equivalent to

$$H_0' : U \sim \mathrm{Ber}(\alpha_0 / (\alpha_0 + \beta_0))$$
$$H_1' : U \sim \mathrm{Ber}(\alpha_1 / (\alpha_1 + \beta_1)). \tag{27}$$

Thus, test statistics for the sequential FHT weighted probability density function can be defined as

$$z_i = \ln[f_0(u_i)/f_1(u_i)]. \tag{28}$$

Then SPRT test statistic defined in Section 2 can be

obtained as

$$z_i = u_i \ln\left(\frac{\alpha_0 \beta_0}{\alpha_1 \beta_1}\right) + \ln\left(\frac{\beta_0(\alpha_1 + \beta_1)}{\beta_1(\alpha_0 + \beta_0)}\right). \quad (29)$$

Hence,

$$\ln L_w = \sum_{i=1}^{n} z_i$$
$$= \sum_{i=1}^{n} u_i \ln\left(\frac{\alpha_0 \beta_0}{\alpha_1 \beta_1}\right) + n \ln\left(\frac{\beta_0(\alpha_1 + \beta_1)}{\beta_1(\alpha_0 + \beta_0)}\right), \quad (30)$$
$$\alpha_j > 0, \beta_j > 0.$$

Expected value of θ_0 and θ_1 is defined as

$$E(z, \theta_0) = \frac{\alpha_0}{\alpha_0 + \beta_0} \ln\left(\frac{\alpha_0 \beta_1}{\alpha_1 \beta_0}\right) + \ln\left(\frac{\beta_0(\alpha_1 + \beta_1)}{\beta_1(\alpha_0 + \beta_0)}\right) \quad (31)$$

$$E(z, \theta_1) = \frac{\alpha_1}{\alpha_1 + \beta_1} \ln\left(\frac{\alpha_0 \beta_1}{\alpha_1 \beta_0}\right) + \ln\left(\frac{\beta_0(\alpha_1 + \beta_1)}{\beta_1(\alpha_0 + \beta_0)}\right). \quad (32)$$

Let N be the random variable denoting the sample size of the fuzzy SPRT. The fuzzy SPRT with α and β, the expected sample size when H_0 is true,

$$E(N/H_0) \approx \frac{(1-\alpha)\ln B + \alpha \ln A}{E(z, \theta)}. \quad (33)$$

Otherwise, when H_1 is true,

$$E(N/H_1) \approx \frac{\beta \ln B + (1-\beta)\ln A}{E(z; \theta)}. \quad (34)$$

As it is mention in Section 2, SPRT's observation equation which is $E(N; \rho) = 2E(N/H_i) + 2, i = 0,1$ is still valid for the FHT of correlation coefficient.

Therefore in this study we showed that fuzzy hypothesis testing for correlation coefficient can be applied using Torabi and Mirhosseini's [9] SPRT.

5. Example

In this section we present an example to clarify the method so far. Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be a sequence from iid r.v.'s from Bivariate Normal distribution. We want to test

$$H_0 : \rho = 0$$
$$H_1 : \rho = 0.80 \quad (35)$$

It is possible to write this hypothesis as follows according to the information given with Theorem 1 in Section 2

$$H_0 : \theta = 0.5$$
$$H_1 : \theta = 0.9 \quad (36)$$

Let $\alpha = 0.05$ and $\beta = 0.1$ then we calculate $E(N/H_0 \text{ is true}) \approx 3.904$ and $(N/H_1 \text{ is true}) \approx 6.4557$.

It is known that Prathan-Sathe's [4] SPRT needs two pair of observations each step. Therefore the expected sample size of this SPRT is given by $E(N, \rho) = 2E(N/H_0 \text{ is true}) + 2 \approx 9.808$. These results are obtained by applying classical hypothesis testing procedure which is assumed that both the underlying hypotheses are crisp. However, we want to test if the correlation coefficient is nearly zero or nearly 0.8. Because fuzzy approaches are extremely popular in real life situations and it ensures to obtain closer values to real results [6,8,11]. Hence crisp hypothesis turns into fuzzy hypothesis given below

$$H_0 : \rho \approx 0$$
$$H_1 : \rho \approx 0.80. \quad (37)$$

It is known that $\alpha = 0.05$ and $\beta = 0.1$. Then we write this hypothesis with Equation 38 by combining the information given in Section 2 and Section 3 as

$$H_0 : \theta \approx 0.5$$
$$H_1 : \theta \approx 0.9. \quad (38)$$

In addition, we determined $\alpha_0 = 0.5$, $\alpha_1 = 0.9$, $\beta_0 = 0.5$, $\beta_1 = 0.1$. Then, membership functions can also be defined as

$$H_j(\theta) = \theta^{\alpha_j - 1}(1-\theta)^{\beta_j - 1}, \forall \theta \in (0,1),$$
$$\theta_j = \alpha_j / (\alpha_j + \beta_j), j = 0,1. \quad (39)$$

Afterwards we calculate $\ln A = -2.89037$ and $\ln B = 2.25129$, $E(Z/H_0 \text{ is true}) = 0.51083$ and $E(Z/H_1 \text{ is true}) = -0.3681$. Hence we obtain $E(N/H_0 \text{ is true}) = 3.9039$ and $E(N/H_1 \text{ is true}) = 6.4553$.

6. Conclusion

In this study, the use of SPRT to test fuzzy hypothesis concerning the correlation coefficient for the bivariate normal distribution is examined. It is known that one can develop SPRT for simple fuzzy hypothesis for correlation coefficient. However, in real world, we may confront with fuzzy hypothesis. According to recent studies, the methods taking into account the statistical hypothesis with fuzzy approaches can be so worthwhile. Thus, in this study, we showed that fuzzy hypothesis testing procedure can be used for SPRT of correlation coefficient.

REFERENCES

[1] C. J. Kowalski, "The OC and ASN Functions of Some SPRT's for the Correlation Coefficient," *Technometrics*, Vol. 13, No. 4, 1971, pp. 833-840. doi:10.1080/00401706.1971.10488853

[2] A. Wald, "Sequential Analysis," John Wiley & Sons, Inc.,

New York, 1947.

[3] S. C. Choi, "Sequential Test for Correlation Coefficients," *Journal of the American Statistical Association*, Vol. 66, No. 335, 1971, pp. 575-576. [doi:10.1080/01621459.1971.10482308](doi:10.1080/01621459.1971.10482308)

[4] M. Pradhan and Y. S. Sathe, "An Unbiased Estimator and a Sequential Test for the Correlation Coefficient," *Journal of the American Statistical Association*, Vol. 70, No. 349, 1975, pp. 160-161. [doi:10.1080/01621459.1975.10480278](doi:10.1080/01621459.1975.10480278)

[5] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, Vol. 8, No. 3, 1965, pp. 338-353. [doi:10.1016/S0019-9958(65)90241-X](doi:10.1016/S0019-9958(65)90241-X)

[6] S. M. Taheri and J. Behboodian, "Neyman-Pearson Lemma for Fuzzy Hypotheses Testing," *Metrika*, Vol. 64, No. 3, 2006, pp. 289-304. [doi:10.1007/s00184-006-0049-8](doi:10.1007/s00184-006-0049-8)

[7] H. Torabi and J. Behboodian, "Sequential Probability Ratio Test for Fuzzy Hypotheses Testing with Vague Data," *Austrian Journal of Statistics*, Vol. 34, No. 1, 2005, pp. 25-38.

[8] R. Talukdar and H. K. Baruah, "Sequential Probability Ratio Test with Fuzzy Observations," *The Journal of Fuzzy Mathematics*, Vol. 15, No. 3, 2007, pp. 627-643.

[9] H. Torabi and S. M. Mirhosseini, "Sequential Probability Ratio Tests for Fuzzy Hypotheses Testing," *Applied Mathematical Sciences*, Vol. 3, No. 33, 2009, pp. 1609-1618.

[10] M. G. Akbari, "Sequential Test of Fuzzy Hypothesis," *American Journal of Statistics*, Vol. 1, No. 2, 2011, pp. 87-92. [doi:10.4236/ojs.2011.12010](doi:10.4236/ojs.2011.12010)

[11] C. H. Leu, C. C. Chen, and C. M. Lai, "Testing Hypothesis Concerning Correlation Coefficient with Fuzzy Data," *5th International Conference on Fuzzy Systems and Knowledge Discovery*, Jinan, 18-20 October 2008, pp. 635-643. doi:10.1109/FSKD.2008.352

[12] H. Torabi and J. Behboodian, "Likelihood Ratio Tests for Fuzzy Hypothesis Testing," *Statistical Papers*, Vol. 48, No. 3, 2007, pp. 509-522. [doi:10.1007/s00362-006-0352-5](doi:10.1007/s00362-006-0352-5)