

## USING FUZZY KNOWLEDGE OF A NUISANCE PARAMETER FOR HYPOTHESIS TESTING\*

J. BEHBOODIAN<sup>1\*\*</sup> AND A. MOHAMMADPOUR<sup>2</sup>

<sup>1</sup>Department of Mathematics, Shiraz Islamic Azad University, Shiraz, I. R. of Iran  
Email: Behboodian@stat.susc.ac.ir

<sup>2</sup>Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology,  
424 Hafez Ave., Tehran, I. R. of Iran, 15914, Present address: L2S, CNRS-Supélec-Univ. Paris 11, France  
Email: adel@aut.ac.ir

**Abstract** – The problem of hypothesis testing with a nuisance parameter is considered. Two methods for using fuzzy knowledge on the nuisance parameter to test hypotheses are suggested. These methods are neither a pure classical nor a pure Bayesian approach to hypothesis testing, but rather related to both. A few known examples and their applications, which cannot be studied by the parametric statistical methods, are discussed.

**Keywords** – Fuzzy Statistics, defuzzified distribution, weighted defuzzification-Estimator, exchangeable normal Distribution

### 1. INTRODUCTION

There are two main approaches to hypothesis testing for fixed sample size: (I) Classical, (II) Bayesian. The classical approach is applied when the hypotheses, observations and parameters are crisp, whereas the Bayesian approach is applied when the hypotheses and observations are crisp, but the parameters are crisp random variables.

There is also a third approach for some practical problems. In this approach, similar to the above two approaches, the hypotheses and observations are crisp, but there be may some fuzzy knowledge about an unknown, but fixed parameter that can be expressed by a known membership function say,  $m(\cdot)$ . This fuzzy knowledge comes from restrictions on the parameter or from our experience. For simplicity, we call such a parameter a fuzzy parameter in this article. Ralescu [1] uses a similar terminology for parameters in binomial distribution and Haekwan & Tanaka [2] introduced regression with fuzzy parameters. This fuzzy knowledge is sometimes quite helpful, but it may not be economical in some problems. To make our point clear, we give the following two examples.

**Example 1. 1.** Suppose we are interested in evaluating the diameters of washers produced by a factory, and we know that the distribution of the difference of such diameters from a norm diameter is normal  $N(\mu, \sigma^2)$ , where  $\sigma^2$  is known. That is, we want to test the hypotheses

---

\*Received by the editor May 22, 2004 and in final revised form September 27, 2005

\*\*Corresponding author

$$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu \neq 0 \end{cases}, \tag{1}$$

based on the random sample  $\mathbf{X} = (X_{1,L}, \dots, X_n)'$  of differences. It is well known that the following test function is the best at level  $\alpha$  when  $\sigma^2$  is known (it is a uniformly most powerful unbiased test (UMPUT)):

$$\phi_1(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\bar{x}}{\sigma/\sqrt{n}} \right| \geq z_{1-\frac{\alpha}{2}} \text{ reject } H_0 \\ 0 & \left| \frac{\bar{x}}{\sigma/\sqrt{n}} \right| < z_{1-\frac{\alpha}{2}} \text{ accept } H_0 \end{cases},$$

where  $\mathbf{x} = (x_{1,L}, \dots, x_n)'$  is an observation vector from  $\mathbf{X}$ ,  $\bar{x} = \sum_{i=1}^n x_i/n$  and  $Z : N(0,1)$  with  $P(Z > z_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$ .

Now suppose  $\sigma^2$  is a fuzzy parameter with membership function  $m(\cdot)$ . How can we test (1) in this case? A simple and logical answer to this question is the usual  $t$ -test, with the following test function:

$$\phi_2(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\bar{x}}{s/\sqrt{n}} \right| \geq t(n-1, 1-\frac{\alpha}{2}) \\ 0 & \left| \frac{\bar{x}}{s/\sqrt{n}} \right| < t(n-1, 1-\frac{\alpha}{2}) \end{cases},$$

where  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ ,  $T$  has a  $t$ -distribution with  $n-1$  degrees of freedom, and  $P(T > t(n-1, 1-\frac{\alpha}{2})) = \frac{\alpha}{2}$ . This test function does not depend on  $\sigma^2$ , but is the best by the generalized likelihood ratio method. It is obvious that the fuzzy knowledge of  $\sigma^2$  is not helpful in this case, since we have the best test function. We can easily show that the power functions of  $\phi_1$  and  $\phi_2$  are close to each other for a reasonable sample size.

**Example 1. 2.** In practical problems, the sample in Example 1.1 is without replacement and so  $\mathbf{X} = (X_{1,L}, \dots, X_n)'$  is a random vector with an exchangeable normal  $EN_n(\mu, \sigma^2, \rho)$  distribution [3], where  $\rho$  is the correlation parameter (see the Appendix). Consider testing (1). It can be shown that the following test functions are the best at level  $\alpha$  when  $\rho$  is known, [4]:

$$\phi_3(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\sqrt{n}\bar{x}}{\sigma\sqrt{1+(n-1)\rho}} \right| \geq z_{1-\frac{\alpha}{2}} \\ 0 & \left| \frac{\sqrt{n}\bar{x}}{\sigma\sqrt{1+(n-1)\rho}} \right| < z_{1-\frac{\alpha}{2}} \end{cases}, \text{ for known } \sigma^2,$$

$$\phi_4(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\sqrt{n(1-\rho)}\bar{x}}{s\sqrt{1+(n-1)\rho}} \right| \geq t(n-1, 1-\frac{\alpha}{2}) \\ 0 & \left| \frac{\sqrt{n(1-\rho)}\bar{x}}{s\sqrt{1+(n-1)\rho}} \right| < t(n-1, 1-\frac{\alpha}{2}) \end{cases}, \text{ for unknown } \sigma^2.$$

If  $\rho$  is unknown there is no known parametric test for testing (1) [4, 5]. However, when we have some fuzzy knowledge about  $\rho$ , we can find a solution for the problem (see Section 3). For example if we know that  $\rho \in [0,1)$  we can classify this restriction in the following membership function,

$$m(\rho) = \begin{cases} 1 & \rho \in [0,1) \\ 0 & \rho \notin [0,1) \end{cases},$$

or, if we know that  $\rho$  is approximately 0.5 (by our experience) we can use the membership functions in Example 2.4 (Fig. 3).

We shall point out the difference between Bayesian and fuzzy knowledge. In the Bayesian framework we consider our parameter as a random variable with a known prior density, which is not a fixed number. But in a fuzzy case, we have only limited fuzzy knowledge about our fixed unknown parameter.

The main purpose of this paper is to show how we can use this fuzzy knowledge for testing hypotheses. However, as far as the authors know, there is no work on this subject i.e., hypothesis testing in classical statistics framework for a crisp parameter when we have some fuzzy knowledge about an unknown fixed nuisance parameter. But some authors work on testing fuzzy hypotheses and hypothesis testing with fuzzy observations. We classify these works in Table 1 of the Appendix.

The distribution function of a random variable which is dependent on a fuzzy parameter (defuzzified distribution function) is defined in the next section. An estimator (weighted defuzzification-estimator) for a fuzzy parameter is also introduced in Section 2. In Section 3 we suggest two methods for testing crisp hypotheses when we have a fuzzy nuisance parameter by using a defuzzified distribution function and weighted defuzzification-estimator. Some advantages and disadvantages of the suggested methods are given in Section 3. Applications of the examples in Section 3 are given in Section 4. The Appendix contains a definition of exchangeable normal distribution and a proof for a theorem which is used in Section 3.

For comparing estimators or test functions we use a Monte Carlo simulation procedure. All computations and plots are done using the S-PLUS software system [6].

## 2. DEFUZZIFICATION

Let  $X$  be a continuous random variable with distribution function  $F_X(x;v)$ , which depends on a fuzzy parameter  $v$  with the known, continuous and integrable membership function  $m(\cdot)$  with the support  $S = \{v \mid m(v) > 0\}$ .

### 2.1. Defuzzified distribution function

Without loss of generality, we assume that the membership function  $m$  is normalized, i.e.  $\int_S m(v)dv = 1$ .

**Definition 2.1.** Let  $X$  have a distribution function depending on a fuzzy parameter  $v$ , where  $v$  has a membership function  $m$ . The *defuzzified distribution function*, *DDF*, of  $X$ ,  $F_X^{\otimes}(x)$ , is defined as the mean or median of  $F_X(x;v)$  over  $m$ .

If we use the mean, to defuzzify  $F_X(x;v)$ , then

$$F_X^{\otimes}(x) = \int_S F_X(x;v) m(v)dv, \quad (2)$$

i.e.  $f_X^{\otimes}$  is a weighted mean of  $F_X(x; \nu)$ . On the other hand, in the Bayesian framework, i.e. when  $\nu$  is a continuous random variable with the prior density  $m$  and  $F_{X|\nu}(x|\nu) = F_X(x; \nu)$  is the conditional distribution function of  $X$ , given  $\nu$  we can calculate the marginal distribution function  $X$ ,  $F_X$  by (2). Therefore we use the notation of  $F_X$  instead of  $f_X^{\otimes}$  in (2).

To simplify calculations of  $f_X^{\otimes}$ , we use the definition of median in statistics. That is, we consider  $\nu$  as a random variable with density function  $m$ , and calculate  $f_X^{\otimes}(x)$  by solving the following equation, using the distribution  $F_X(x; \nu)$ ,

$$F_{F_X(x; \nu)}(f_X^{\otimes}(x)) = \frac{1}{2} \text{ or } P(F_X(x; \nu) \leq f_X^{\otimes}(x)) = \frac{1}{2}. \quad (3)$$

The following theorem states an important property of  $f_X^{\otimes}$ .

**Theorem 2. 1.**  $f_X^{\otimes}$  is a non-decreasing function.

**Proof:** Let  $x_1 < x_2$ . For  $i = 1, 2$ , take  $k_i = f_X^{\otimes}(x_i)$ ,  $Y_i = F_X(x_i; \nu)$ , and so

$$P(Y_1 \leq k_1) = P(Y_2 \leq k_2) = \frac{1}{2}, \text{ and } Y_1 \leq Y_2.$$

Therefore,

$$P(Y_1 \leq k_1) = P(Y_2 \leq k_2) \leq P(Y_1 \leq k_2),$$

i.e.  $k_1 \leq k_2$  or equivalently  $f_X^{\otimes}$  is non-decreasing.

If  $f_X^{\otimes}$  is a non-decreasing function, then  $f_X^{\otimes}(x-) = \lim_{t \uparrow x} f_X^{\otimes}(t)$ ,  $f_X^{\otimes}(x+) = \lim_{t \downarrow x} f_X^{\otimes}(t)$  exist and are finite. Further,  $F_X(x; \nu)$  is continuous with respect to  $x$ , and so

$$P(F_X(x-; \nu) \leq f_X^{\otimes}(x-)) = P(F_X(x; \nu) \leq f_X^{\otimes}(x-)),$$

$$P(F_X(x+; \nu) \leq f_X^{\otimes}(x+)) = P(F_X(x; \nu) \leq f_X^{\otimes}(x+)).$$

And by (3) we have

$$P(F_X(x; \nu) \leq f_X^{\otimes}(x-)) = P(F_X(x; \nu) \leq f_X^{\otimes}(x)) = P(F_X(x; \nu) \leq f_X^{\otimes}(x+)). \quad (4)$$

If  $Y = F_X(x; \nu)$  has a unique median, then  $f_X^{\otimes}(x-) = f_X^{\otimes}(x) = f_X^{\otimes}(x+)$  by (2.3) i.e.  $f_X^{\otimes}(x)$  is continuous.

On the other hand,  $f_X^{\otimes}(x)$  is the median of random variable  $0 \leq Y \leq 1$ , and so  $0 \leq f_X^{\otimes}(x) \leq 1$ . Also, by Theorem 2.1,  $f_X^{\otimes}(+\infty)$  and  $f_X^{\otimes}(-\infty)$  exist as  $\lim_{t \uparrow +\infty} f_X^{\otimes}(t)$ , and  $\lim_{t \downarrow -\infty} f_X^{\otimes}(t)$  respectively, [7]. Therefore  $f_X^{\otimes}(x)$  is a distribution function if  $f_X^{\otimes}(+\infty) = 1$  and  $f_X^{\otimes}(-\infty) = 0$ , such as the following two examples.

**Example 2. 1.** Let  $X$  be exponentially distributed, i.e.

$$F_X(x; \nu) = 1 - \exp(-\nu x), \quad x > 0,$$

where  $m(\nu) = 1$ ,  $0 < \nu \leq 1$ . In this example we can calculate  $f_X^{\otimes}$  exactly by (3) as follows

$$\begin{aligned}
 P(1 - \exp(-vx) \leq f_x^{\otimes}(x)) &= \frac{1}{2} \\
 \iff P(v \leq \frac{-1}{x} \ln(1 - f_x^{\otimes}(x))) &= \frac{1}{2} \\
 \iff \frac{-1}{x} \ln(1 - f_x^{\otimes}(x)) &= \frac{1}{2} \\
 \iff f_x^{\otimes}(x) &= 1 - \exp(-x/2).
 \end{aligned}$$

Moreover,

$$F_x(x) = 1 + \frac{1}{x}(\exp(-x) - 1), x > 0.$$

**Example 2.2.** Suppose that  $m(v) = \exp(-v), v > 0$  in Example 2.1. In this case,

$$f_x^{\otimes}(x) = 1 - \exp(x \ln(1/2)), x > 0,$$

$$F_x(x) = 1 - \frac{1}{x+1}, x > 0.$$

Numerical methods for calculating DDFs are introduced in [8]. In the next section we assume that  $F^{\otimes}$  is a distribution function.

## 2.2. Weighted defuzzification-estimator

We want to estimate the fuzzy parameter by an observation vector  $\mathbf{x}$ . In classical statistics there are several methods for finding estimators, but in this problem, besides the observation vector  $\mathbf{x}$ , we can use the fuzzy knowledge about  $v$  to find a better estimator. (This is similar to Bayesian estimation though  $v$  is not a random variable.) We call such an estimator a weighted defuzzification-estimator (WDE) for a fuzzy parameter.

The first step to calculate a WDE for  $v$  is to defuzzify  $v$ , i.e. we convert the fuzzy knowledge about  $v$  to a crisp number. There are several ways to accomplish this [9]. For example, we can use the mean or median of the membership function (MoM) for defuzzification of  $v$ , say  $v_{\otimes}$ .

The second step is to estimate  $v$  by an observation vector  $\mathbf{x}$  on the restricted parameter space, i.e.  $S$ . For example, we can use the maximum likelihood estimator (MLE) of  $v$ , say  $\hat{v}$ . The final step is combining the results of the two previous steps. We can use a weighted mean for combining  $v_{\otimes}$  and  $\hat{v}$  as follows:

$$v_w = w v_{\otimes} + (1-w) \hat{v}, \quad 0 \leq w \leq 1, \quad (5)$$

and one of the intuitive choices for the weight  $w$  is

$$w = \frac{m(v_{\otimes})}{m(v_{\otimes}) + m(\hat{v})}, \quad 1-w = \frac{m(\hat{v})}{m(v_{\otimes}) + m(\hat{v})}.$$

We call  $v_w$  a WDE for the fuzzy parameter  $v$ . Therefore we can define the WDE as follows:

**Definition 2.2.** A WDE for a fuzzy parameter is a convex combination of a defuzzification and an estimator for this fuzzy parameter such as (5).

There are some interesting properties of the WDE's as follow, which can be easily shown:

1. If  $S = \{v\}$ , i.e.  $v$  is a known parameter, then  $v_w = v$ .
2. If  $v_w = v$ , and  $E\hat{v} = v$  then  $Ev_w = v$ , i.e. WDE is an unbiased estimator for  $v$ .
3. If  $m(v_w) \neq 0$  then  $Var(v_w) < Var(\hat{v})$ .
4. WDE is a range preserver estimator, i.e.  $v_w \in S$ .
5. WDE is a sufficient statistic for  $v$ .
6. WDE depends on the sample only through a minimal sufficient statistic.
7. WDE does not depend on the sampling plan.
8. WDE attains the Cramer-Rao lower bound for estimating its expectation.

We recall that a MLE has the properties 4-8, and a Bayes estimator has property 6, [10].

In the following examples we compare the mean squared error (MSE) of a classical estimator with a WDE.

**Example 2. 3.** Let  $\mathbf{X} : EN_n(\mu, 1, \rho)$ , where  $\rho$  is unknown and  $\mu$  is a fuzzy parameter. Consider the following membership functions:

$$m_l(\mu) = \begin{cases} 0 & \mu < -1 \\ 1 + \mu & -1 \leq \mu < 0 \\ 1 - \mu & 0 \leq \mu < 1 \\ 0 & 1 \leq \mu \end{cases},$$

$$m_p(\mu) = \begin{cases} 0 & \mu < -1 \\ 1 - \mu^2 & -1 \leq \mu < 1 \\ 0 & 1 \leq \mu \end{cases},$$

$$m_h(\mu) = \begin{cases} 0 & \mu < -1 \\ \frac{-\mu - 1}{\mu - 1} & -1 \leq \mu < 0 \\ \frac{-\mu + 1}{\mu + 1} & 0 \leq \mu < 1 \\ 0 & 1 \leq \mu \end{cases},$$

which are linear, parabolic, and hyperbolic respectively with zero MoM (see Fig. 1).

If we calculate the WDE of  $\mu$  by using the MoM and MLE we have

$$\mu_{w_j} = (1 - w_j) \bar{X}, \quad w_j = \frac{1}{1 + m_j(\bar{X})}, \quad j = l, p, h.$$

It is easy to prove that the MSE for  $\bar{X}$  is given by  $\frac{1+(n-1)\rho}{n}$ . Calculation of the exact MSE for WDEs is not easy, but it can be approximated. Figure 2 shows the MSEs for WDEs with  $m_l$ ,  $m_p$  and  $m_h$  as functions of  $\mu$  for  $\rho = 0.2, 0.8$ , and different sample sizes. We also plot the exact MSE of  $\bar{X}$  in Fig. 2. We see that when  $\rho$  is large, the WDEs for  $\mu$  are much better than  $\bar{X}$ , and the differences among WDEs are due to the shapes of the membership functions.

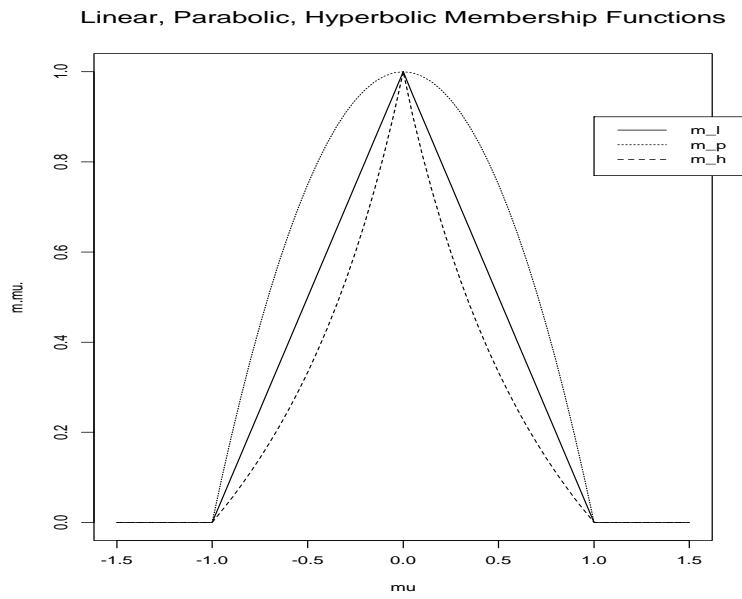


Fig. 1. Graphs of  $m_i(\mu), i = l, p, h$  in Example 2.3

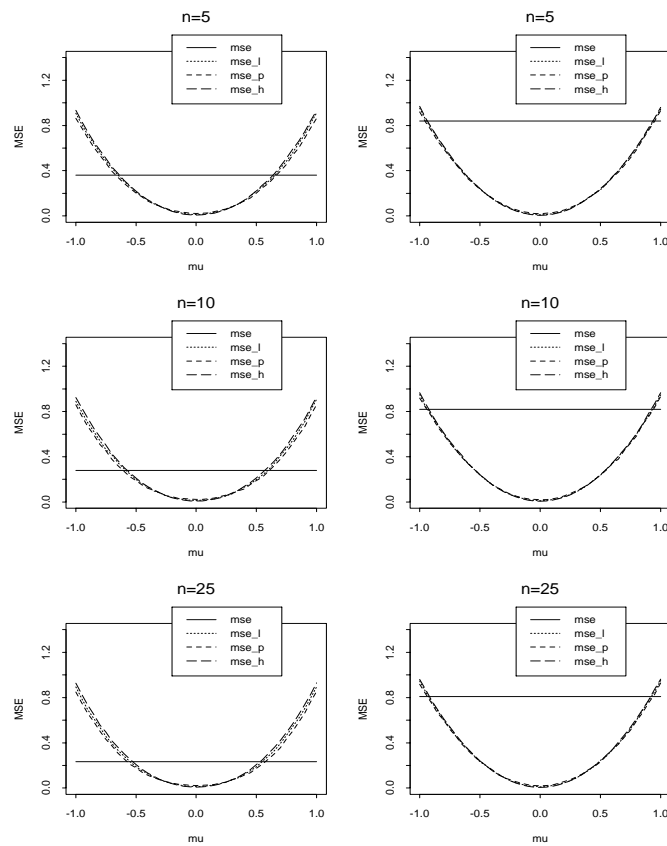


Fig. 2. The MSE for MLE and WDEs in Example 2.3, left:  $\rho = 0.2$ , right:  $\rho = 0.8$ . We use the following abbreviation: mse: MSE of the MLE; mse<sub>j</sub>: MSE of the WDE when we use the membership function  $m_j(\cdot), j = l, p, h$

**Example 2. 4.** Let  $X : EN_n(\mu, 1, \rho)$ , where  $\mu$  is unknown and  $\rho$  is a fuzzy parameter. Consider the following membership function  $s$  (Fig. 3):

$$m_i(\rho) = \begin{cases} \frac{\rho - \frac{i-1}{6}}{\frac{1}{2} - \frac{i-1}{6}} & \frac{i-1}{6} \leq \rho < \frac{1}{2} \\ \frac{\rho - \frac{7-i}{6}}{\frac{1}{2} - \frac{7-i}{6}} & \frac{1}{2} \leq \rho < \frac{7-i}{6}, \quad i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$$

Triangular Membership Functions

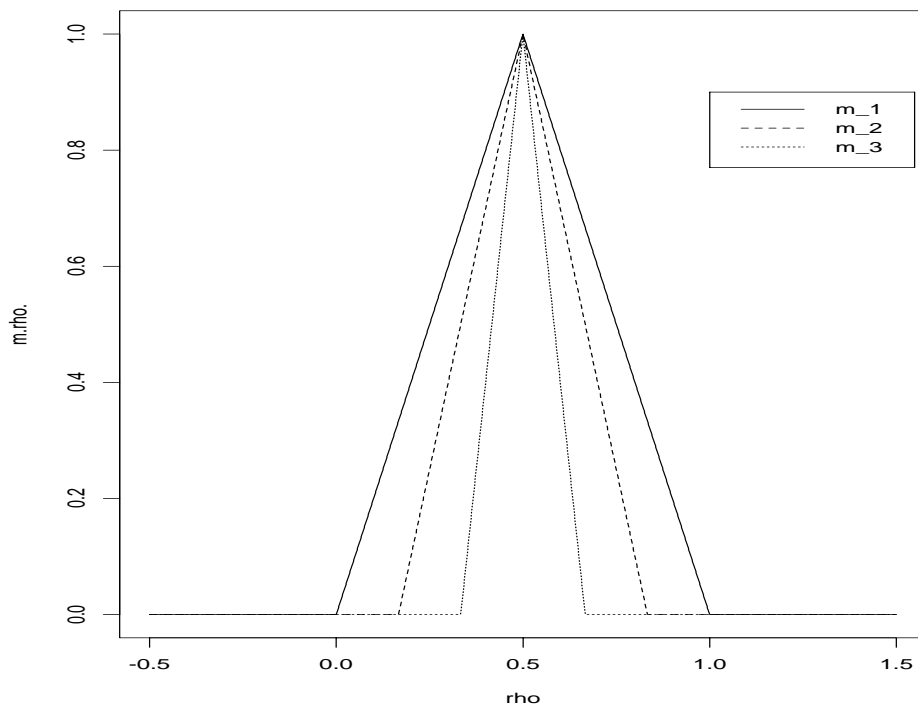


Fig. 3. Graphs of  $m_i(\rho)$ ,  $i = 1, 2, 3$  in Example 2.4

We can approximate the MSE for WDEs and MLE of  $\rho$ . Note that in this case the value of  $\mu$  does not have any effect on MLE or WDEs of  $\rho$ , therefore we assume that  $\mu = 0$ . Figure 4 shows the MSEs for  $n = 5, 25$ . We see that the WDE based on  $m_3$  is better than the other WDEs and MLE.



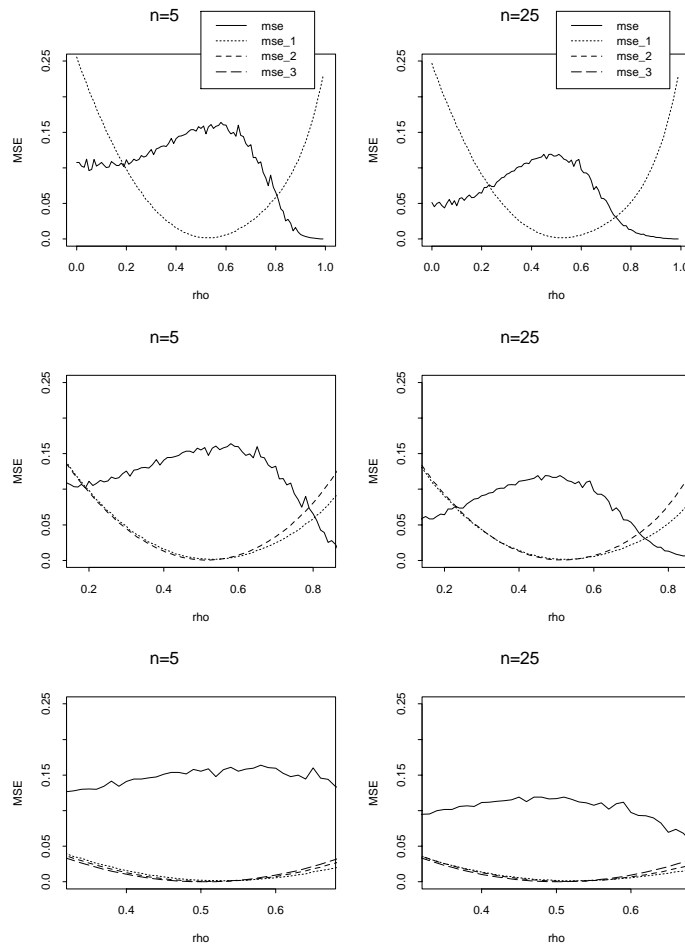


Fig. 4. The MSE for MLE and WDEs in Example 2.4. We use the following abbreviation: mse: MSE of the MLE; mse\_i: MSE of the WDE when we use the membership function  $m_i(\cdot)$ ,  $i = 1, 2, 3$

### 3. HYPOTHESIS TESTING

Let  $\mathbf{x}$  be an  $n$ -dimensional observation vector from a distribution with two unknown parameters,  $\theta$  and  $v$ . We want to test the hypotheses

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}, \quad \Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \cup \Theta_1 = \Theta, \quad (6)$$

where  $\Theta$  is the parameter space of  $\theta$ . In some cases, there is no best test function for (6), however when  $v$  is a known parameter (similar to Example 1.2), the main test function can be obtained. We denote this test function by

$$\phi_v(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in C_v, \\ 0 & \mathbf{x} \notin C_v, \end{cases} \quad (7)$$

where  $C_v$  is the critical region for a given value of  $v$ . The type I error is given by

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta(\phi_v(\mathbf{X}) = 1) = P_{\theta_0}(\phi_v(\mathbf{X}) = 1), \theta_0 \in \overline{\Theta_0},$$

where  $\overline{\Theta_0}$  is the boundary of  $\Theta_0$  and  $P_\theta$  is the probability function with parameter  $\theta$ .

Consider the cases where there is no best test function for (6). In this section we answer the following question:

How can we use the fuzzy knowledge about the nuisance parameter  $v$  for testing hypotheses (6)?

If  $v$  is an unknown parameter then we cannot make a decision by test function (7) for testing hypotheses (6). We suggest the following procedure for testing (6). If we calculate a WDE for  $v$ , i.e.,  $v_w$  and insert it in  $\phi_v$ , then we have a test function  $\phi_{v_w}$ , which does not depend on the unknown parameter. But the size of this test may not be equal to  $\alpha$ . Hence we should find a new critical region  $C'_{v_w}$  such that

$$P_{\theta_0}(\phi_{v_w}(\mathbf{X}) = 1; v) = \alpha, \text{ and } \phi_{v_w}(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in C'_{v_w}, \\ 0 & \mathbf{x} \notin C'_{v_w}, \end{cases} \tag{8}$$

where (8) should be calculated by the DDF of a function of  $\mathbf{X}$ . In the following examples we describe this method of hypothesis testing.

**Example 3.1.** Suppose we want to test

$$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu \neq 0 \end{cases},$$

in Example 2.4. The best test is given by  $\phi_3$  in Example 1.2, when  $\rho$  is a crisp known parameter. If  $\rho$  is a fuzzy parameter with membership function  $m_1$ , then the test function with our suggested method is given by

$$\phi_{3,1}(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\sqrt{n}\bar{x}}{\sqrt{1+(n-1)\rho_{w_1}}} \right| \geq k_{n,1} \\ 0 & \left| \frac{\sqrt{n}\bar{x}}{\sqrt{1+(n-1)\rho_{w_1}}} \right| < k_{n,1} \end{cases},$$

where  $\rho_{w_1}$  is the WDE of  $\rho$  based on membership function  $m_1$ , and  $k_{n,1}$  can be approximated by

$$F_Y(k_{n,1}) = 1 - \alpha, Y = \left| \frac{\sqrt{n}\bar{X}}{\sqrt{1+(n-1)\rho_{w_1}}} \right|,$$

(DDF of  $Y$  based on mean). Figure 5 shows the exact power function for  $\phi_3$  with  $\rho = 0.5$  and the approximated power function for  $\phi_{3,1}$ , when  $\alpha = 0.05$ . Note that  $\phi_{3,1}$  is very close to  $\phi_3$ . This is due to the fact that MLE, and hence WDE, does not depend on  $\mu$  and the WDE of  $\rho$  is very close to the real value of  $\rho$ .

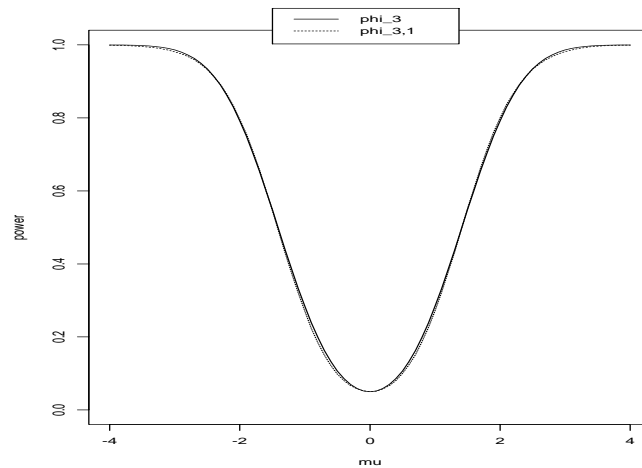


Fig. 5. The power functions for  $\phi_3$ , and  $\phi_{3,1}$  for  $n = 25$ , when  $\alpha = 0.05$ . We use the following abbreviation:

phi\_3: power function of  $\phi_3$ , when  $\rho = 0.5$ ; phi\_3,1: power function of  $\phi_{3,1}$

**Example 3. 2.** Let  $\mathbf{X} : EN_n(\mu, \sigma^2, \rho)$ , where  $\sigma^2$  and  $\rho$  are unknown crisp parameters, and  $\mu$  is a fuzzy parameter. Consider the membership functions in Example 2.3, for  $\mu$ . We want to test

$$\begin{cases} H_0 : \rho = 0, \\ H_1 : \rho > 0. \end{cases} \quad (9)$$

The following test function is the best when  $\mu$  is known (see the Appendix):

$$\phi_\mu(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| \geq t(n-1, 1 - \frac{\alpha}{2}), \\ 0 & \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| < t(n-1, 1 - \frac{\alpha}{2}). \end{cases}$$

The test functions with our suggested method are given by

$$\phi_{\mu_{w_j}}(\mathbf{x}) = \begin{cases} 1 & \left| \frac{\bar{x} - \mu_{w_j}}{s/\sqrt{n}} \right| \geq k_{n,j} \\ 0 & \left| \frac{\bar{x} - \mu_{w_j}}{s/\sqrt{n}} \right| < k_{n,j}, \end{cases} \quad j = l, p, h,$$

where  $\mu_{w_j}$  is the WDE for  $\mu$  by using membership functions,  $m_j$ ,  $j = l, p, h$ , and  $k_{n,j}$  can be approximated by

$$F_{Y_j}(k_{n,j}) = 1 - \alpha, \text{ or } \tilde{F}_{Y_j}(k_{n,j}) = 1 - \alpha, Y_j = \left| \frac{\bar{X} - \mu_{w_j}}{S/\sqrt{n}} \right|, j = l, p, h,$$

(DDF based on mean or median respectively). We approximate the critical values for  $\sigma^2 = 1$  and  $\alpha = 0.05$ . Figure 6 shows the power functions of  $\phi_{\mu_{w_j}}$ ,  $j = l, p, h$ , (calculated by using  $F_{Y_j}$ ) for

different sample sizes. In Fig. 7 we plot the power function  $\phi_{\mu_{w_l}}$  based on  $F_{Y_l}$  and  $F_{Y_l}^{\alpha}$ . We also plot the exact power function of  $\phi_{\mu}$  when  $\mu = 0$  in Figs. 6 and 7.

In the case that  $n = 2$  (Fig. 7) the calculated power function based on  $F_{Y_l}^{\alpha}$  dominate the calculated power function based on  $F_{Y_l}$  for  $\rho > 0.2$ . This is due to the fact that the distribution of the test function is heavy tailed (when  $\mu$  is known the test function has Cauchy distribution) and the median is robust with respect to outlier probabilities.

Note that if  $\mu$  is unknown, there is no best test function for testing (9). (See the Appendix, and [3].) If

$$m(\mu) = \begin{cases} 1 & \mu = 0 \\ 0 & \mu \neq 0 \end{cases}$$

then the test function for testing (9) is equal to  $\phi_{\mu}$ , i.e. the suggested method gives the ordinary test function. This is the main reason that the power function of  $\phi_{\mu}$  dominates the other power functions. The power function of  $\phi_{\mu_{w_p}}$  is lower than the other power functions in Fig. 6. This is due to the fact that, the vagueness of  $m_p$  (about  $\mu$ ) is more than the other membership functions. In Fig. 6 we show that

$$P_{\rho}(\phi_{\mu_{w_p}}(\mathbf{X}) = 1; \mu) \leq P_{\rho}(\phi_{\mu_{w_l}}(\mathbf{X}) = 1; \mu) \leq P_{\rho}(\phi_{\mu_{w_h}}(\mathbf{X}) = 1; \mu) \leq P_{\rho}(\phi_{\mu}(\mathbf{X}) = 1; \mu = 0).$$

Of course, it is not easy to prove this result. But we can prove the following theorem.

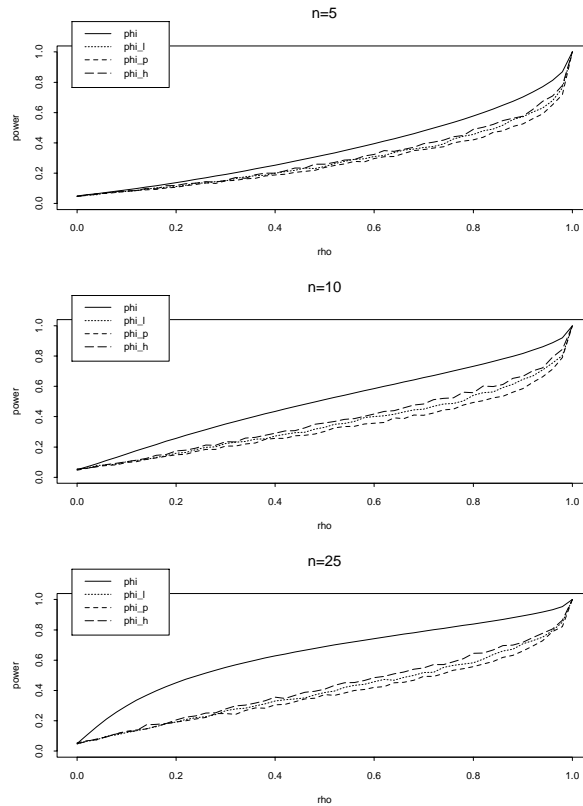


Fig.6. The power functions for  $\phi_{\mu=0}$ , and  $\phi_{\mu_{w_j}}$ ,  $j = l, p, h$ , for  $n = 5, 10, 25$ , when  $\alpha = 0.05$ . We use the following abbreviation: phi: power function of  $\phi_{\mu=0}$ ; phi\_j: power function of  $\phi_{\mu_{w_j}}$ ,  $j = l, p, h$

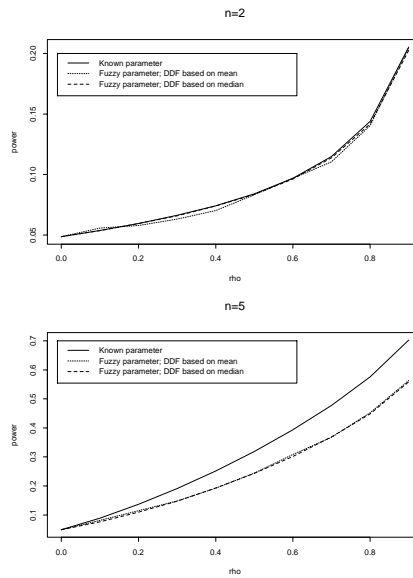


Fig.7. The power functions for  $\phi_{\mu=0}$ , and  $\phi_{\mu_w}$  (by using two methods of DDF), for  $n = 2, 5$ , when  $\alpha = 0.05$

**Theorem 3. 2.** Consider testing (6), based on an  $n$  - dimensional observation vector  $\mathbf{x}$  with density function  $f(\mathbf{x};\nu)$ . The test function (7) is of size  $\alpha$  when  $\nu$  is known, and (8) is of size  $\alpha$  when  $\nu$  is a fuzzy parameter with known membership function  $m$  on the support  $S$ . Suppose there exists  $\nu_0 \in S$  such that  $C'_{\nu_w} \subseteq C_{\nu_0}$  and  $f(\mathbf{x};\nu) \leq f(\mathbf{x};\nu_0), \forall \theta \in \Theta_1, \forall \nu \in S$ . Then the power function of  $\phi_{\nu_0}$  dominates the power function of  $\phi_{\nu_w}$  on  $\Theta_1$ .

**Proof:** We will prove the theorem for the case that  $f_\nu(\mathbf{x})$  and  $m(\nu)$  are continuous functions. The proof for the discrete case can be accomplished by replacing integrals with sums. The power function of  $\phi_{\nu_w}$  is equal to:

$$\begin{aligned}
 P(\phi_{\nu_w}(\mathbf{X})=1; \nu) &= \frac{\int_S P(\phi_{\nu_w}(\mathbf{X})=1) m(\nu) d\nu}{\int_S m(\nu) d\nu} \\
 &= \frac{\int_S \int_{\mathbf{R}^n} \phi_{\nu_w}(\mathbf{x}) f(\mathbf{x};\nu) m(\nu) d\mathbf{x} d\nu}{\int_S m(\nu) d\nu} \\
 &= \frac{\int_{\mathbf{R}^n} \phi_{\nu_w}(\mathbf{x}) \int_S f(\mathbf{x};\nu) m(\nu) d\nu d\mathbf{x}}{\int_S m(\nu) d\nu} \\
 &\leq \frac{\int_{\mathbf{R}^n} \phi_{\nu_w}(\mathbf{x}) \int_S f(\mathbf{x};\nu_0) m(\nu) d\nu d\mathbf{x}}{\int_S m(\nu) d\nu}, \theta \in \Theta_1 \\
 &= \int_{\mathbf{R}^n} \phi_{\nu_w}(\mathbf{x}) f(\mathbf{x};\nu_0) d\mathbf{x}, \theta \in \Theta_1.
 \end{aligned}$$

But  $C'_{\nu_w} \subseteq C_{\nu_0}$  and so  $\phi_{\nu_w}(\mathbf{x}) \leq \phi_{\nu_0}(\mathbf{x}), \forall \mathbf{x} \in \mathbf{R}^n$ . Therefore,

$$P(\phi_{v_w}(\mathbf{X})=1; v) \leq \int_{\mathbb{R}^n} \phi_{v_0}(\mathbf{x}) f(\mathbf{x}; v_0) d\mathbf{x}, \theta \in \Theta_1$$

$$= P(\phi_{v_0}(\mathbf{X})=1), \theta \in \Theta_1.$$

**Example 3.3.** Let  $X : N(\mu, \sigma^2)$ , where  $\sigma^2$  is a fuzzy parameter with membership function

$$m(\sigma^2) = \begin{cases} \frac{2}{3} & \sigma^2 = 1 \\ \frac{1}{3} & \sigma^2 = 4 \\ 0 & \text{otherwise} \end{cases}.$$

Consider testing  $\begin{cases} H_0 : \mu \geq 0 \\ H_1 : \mu < 0 \end{cases}$ . If  $\sigma^2$  is known, then the family of normal distribution has a MLR (Monotone Likelihood Ratio) property (Fig. 8) and according to the Karlin-Robin theorem [11]

$$\phi_{\sigma^2}(x) = \begin{cases} 1 & \frac{x}{\sigma} < z_\alpha \\ 0 & \frac{x}{\sigma} > z_\alpha \end{cases},$$

is the best (UMP) of size  $\alpha$  test function (If  $\sigma^2$  is unknown, then the best test does not exist).  $\phi_{\sigma_w^2}(x) = \begin{cases} 1 & x < k \\ 0 & x > k \end{cases}$  is a test function with our suggested method, where  $\sigma_w^2$  is the WDE for  $\sigma^2$ , and  $k$  can be calculated exactly from  $P(\phi_{\sigma_w^2}(X) = 1; \sigma^2) = \alpha$ . For example  $k = -2.2784$  when  $\alpha = 0.05$ . The assumptions of Theorem 3.2 hold in this example for  $\sigma^2 = 1$ . Therefore, for  $\mu \leq 0$

$$P(\phi_{\sigma_w^2}(X) = 1; \sigma^2) = \frac{2}{3} \Phi(k - \mu) + \frac{1}{3} \Phi\left(\frac{k - \mu}{2}\right)$$

$$\leq P(\phi_{\sigma^2=1}(X) = 1)$$

$$= \Phi(z_\alpha - \mu),$$

where  $\Phi(\cdot)$  is the standard normal distribution function, and  $\Phi(z_\alpha) = \alpha$ . Figure 8 shows these power functions.

Note that if  $\sigma^2$  is a random variable with density  $m$  (and  $X | \sigma^2 : N(\mu, \sigma^2)$ ), then the density of  $X$  is given by

$$f_X(x) = \frac{2}{3} \varphi(x - \mu) + \frac{1}{3} \frac{1}{2} \varphi\left(\frac{x - \mu}{2}\right),$$

where  $\varphi(\cdot)$  is the standard normal density function. In this case the family of  $X$  does not have a MLR property (Fig. 8), thus the Karlin-Robin theorem for finding the UMP test cannot be used. But

for simple hypotheses, for example  $\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu = -1 \end{cases}$ , there exists an MP test. That is, the best test is given by the Neyman-Pearson lemma as follows

$$\phi_{NP}(x) = \begin{cases} 1 & \frac{\frac{2}{3}\varphi(x+1) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x+1}{2})}{\frac{2}{3}\varphi(x) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x}{2})} > k, \\ 0 & \frac{\frac{2}{3}\varphi(x+1) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x+1}{2})}{\frac{2}{3}\varphi(x) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x}{2})} < k, \end{cases}$$

where  $k = 2.94$  for  $\alpha = 0.05$  and its power is equal to 0.157, but with our suggested method, is equal to 0.154. These are the advantages and disadvantages of our methods. But we point out the fact that: In real applied problems the nuisance parameter is either random or non-random, and so we should apply the corresponding methods of hypothesis testing in each case.

The suggested method for hypothesis testing has two restrictions: first, it may not exist (because WDE may not exist), and we cannot warrant its efficiency, except by plotting the power function (because WDE may not be a good estimator). In these cases we suggest the following method for hypothesis testing when we have fuzzy knowledge about a nuisance parameter.

In Section 2.1, we showed that DDF,  $F_{\alpha}^{\mathcal{F}}$ , is a distribution function under a few conditions. If DDF depends on some unknown parameters, we can apply classical methods in statistics to do inference about unknown parameters. We can easily extend the well known theorems such as the Karlin-Robin theorem to the cases when we have fuzzy knowledge about an unknown parameter. The following result states the Neyman-Pearson lemma.

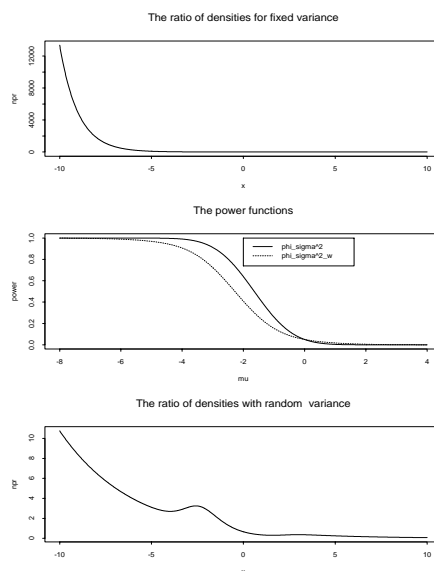


Fig. 8. Top:  $npr(x) = \frac{\varphi(x+1)}{\varphi(x)}$ . Middle: The power functions for  $\phi_{\sigma^2=1}$ , and  $\phi_{\sigma_w^2}$ , when  $\alpha = 0.05$ . We use the following abbreviation: phi\_sigma^2: power function of  $\phi_{\sigma^2=1}$ ; phi\_sigma^2\_w: power

function of  $\phi_{\sigma_w^2}$  Below:  $npr(x) = \frac{\frac{2}{3}\varphi(x+1) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x+1}{2})}{\frac{2}{3}\varphi(x) + \frac{1}{3}\frac{1}{2}\varphi(\frac{x}{2})}$

**Lemma 3. 1.** Consider testing

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta = \theta_1, \end{cases} \tag{10}$$

where  $\theta$  is an unknown parameter of  $f_{\theta}^{\otimes X}$  (DDF of random variable  $X$ ) and  $\theta_0, \theta_1$  are fixed known numbers. If  $f_{\theta}^{\otimes X}$  does not depend on unknown parameters under  $H_0, H_1$ , then

$$\phi(x) = \begin{cases} 1 & \frac{d f_{\theta}^{\otimes X}(x)}{dx} \Big|_{\theta=\theta_1} > k \frac{d f_{\theta}^{\otimes X}(x)}{dx} \Big|_{\theta=\theta_0}, \\ 0 & \frac{d f_{\theta}^{\otimes X}(x)}{dx} \Big|_{\theta=\theta_1} < k \frac{d f_{\theta}^{\otimes X}(x)}{dx} \Big|_{\theta=\theta_0}, \end{cases} \tag{11}$$

for some  $k \geq 0$  is an MP test of its size, say  $\alpha$ , for testing (3.5).

**Proof:** Take  $f_{\theta}^{\otimes X}(x) = \frac{d f_{\theta}^{\otimes X}(x)}{dx}$ .  $f_{\theta}^{\otimes X}(x)$  is a continuous density function which is not dependent on the unknown parameter  $\theta$  under  $H_0, H_1$ . Therefore by the Neyman-Pearson lemma, (11) is an MP test of its size for testing (10) [12].

The following example shows that, the power functions of a test function with fuzzy and random nuisance parameters are different, and the answer in the fuzzy case is much better.

**Example 3. 4.** Let  $X : N(\mu, \sigma^2)$ , where  $\sigma$  is a fuzzy parameter with uniform or exponential membership function (were defined in Examples 2.1 and 2.2, respectively). Consider testing

$$\begin{cases} H_0 : \mu = 0, \\ H_1 : \mu < 0, \end{cases}$$

based on an observation  $x$  from  $X$ . If  $\sigma$  is known, then

$$\phi_{\sigma}(x) = \begin{cases} 1 & x < k, \\ 0 & x > k, \end{cases} \tag{12}$$

is the best test of size  $\alpha$ , where  $k = \sigma z_{\alpha}$ . In the case that  $\sigma$  is a fuzzy parameter,  $k$  should be calculated from  $f_{\theta}^{\otimes X}(k) = \alpha$ , where  $f_{\theta}^{\otimes X}(x)$  is a DDF of  $X$ .

Figure 9 shows the graphs of power functions of the test function (12), when the standard deviation of  $X$  is a fuzzy parameter with a uniform membership function (top) and an exponential membership function (below) for  $\alpha = 0.05$ . We also plot the graphs of the power function  $\phi_{\sigma}$  for  $\sigma = 0.4, 1$  (fixed standard deviation). The graphs show that the test by using DDF based on the median is much better than the test by using DDF based on the mean. The result is interesting, because in the Bayesian framework, i.e., when  $\sigma$  is a random variable with uniform or exponential prior density function, the graph of the power function of (12) is the same as the test function in the fuzzy case when we use DDF based on the mean.



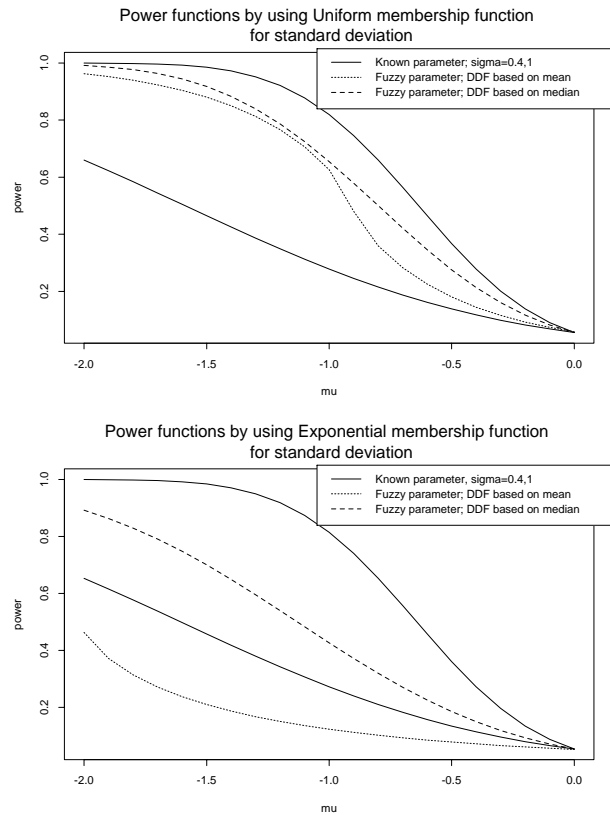


Fig. 9. Graphs of power functions for testing mean,  $\mu$  in a Normal distribution, with fixed, random (fuzzy, DDF based on mean), and fuzzy standard deviation,  $\sigma$ , when  $\alpha = 0.05$

#### 4. APPLICATIONS AND CONCLUSIONS

We have shown in this work that we can use fuzzy knowledge about a nuisance parameter in classical statistics for testing hypotheses. We introduced two methods of hypothesis testing based on a definition for the distribution function of a random variable with a fuzzy parameter (called DDF) and estimation of a fuzzy parameter (called WDE). These methods of hypothesis testing enable us to study some of the problems which cannot be studied in classical statistics by a parametric method. Example 3.1 was concerned with the problem of hypothesis testing for the mean of an exchangeable normal population. This problem was an open problem [5]. In Example 3.2 we presented a parametric solution for testing the independence assumption versus the exchangeability assumption for normal distribution. This problem was also an open problem in linear models [3], time series [13] (for checking error terms), and quality control [13]. Note that, the non-parametric tests for these problems are not robust and cannot be used in many real problems [13, 4].

We showed a few properties of suggested methods in Theorems 3.1, and 3.2. Some advantages and disadvantages of our suggested methods, with respect to the classical and Bayesian framework for hypothesis testing, were given in Examples 3.2, 3.3, and 3.4. Moreover, it is shown that the result of our method was not necessarily the same as other methods.

**Acknowledgments-** The authors would like to thank the referees for their helpful comments. They are also grateful to the Statistical Research Center, Tehran, Iran, for the support of this work.

## REFERENCES

1. Ralescu, D. (1995). *Statistical decision analysis with fuzzy information*. manuscript.
2. Haekwan, L. & Tanaka, H. (1999). Fuzzy approximations with non-symmetric fuzzy parameters in fuzzy regression analysis. *J. Oper. Res. Soc. Japan*, 42, 98-112.
3. Arnold, S. F. (1979). Linear models with exchangeably distributed errors. *Journal of the American Statistical Association*, 74,194-199.
4. Mohammadpour, A. & Behboodian, J. (2000). Hypothesis testing for an exchangeable normal distribution. *J. Sci. I. R. Iran*, 11,131-141.
5. Rao, C. R. (1973). *Linear statistical inference and its applications*, 2nd ed, New York, Wiley.
6. Statistical Sciences, S-PLUS user's manual, version 3.3 StatSci, a division of MathSoft Inc., 1700 Westlake Ave., Suite 500, Seattle (1995). WA 98109-9891.
7. Rohatgi, V. K. (1976). *An Introduction to probability theory and mathematical statistics*. New York, Wiley.
8. Mohammadpour, A. (2003). *Fuzzy parameter and its application in hypothesis testing*. Technical report, School of Intelligent Systems, IPM, Tehran, Iran.
9. Zimmermann, H. J. (1996). *Fuzzy set theory-and its applications*.third ed, Dordrecht, Kluwer Academic Publishers.
10. Migon, H. S. & Gamerman, D. (1999). *Statistical inference: an integrated approach*. London, Arnold.
11. Casella, G. & Berger, R. L. (1990). *Statistical inference*, Brooks/ Cole, Ca.
12. Lehmann, E. L. (1991). *Testing statistical hypotheses*. 2nd ed, California, Wadsworth.
13. Dufour, J. M. & Roy, R. (1986). L'échangeabilité en séries chronologiques: quelques résultats exacts sur les autocorrelations et les statistiques Portemanteau. *Cahiers du C.E.R.O.*, 28, 19-39.
14. Tong, Y. L. (1990). *The multivariate normal distribution*, New York, Springer-Verlag.
15. Ferguson, T. S. (1967). *Mathematical statistics: a decision theoretic approach*. New York, Academic Press.
16. Fisher, R. A. (1925). *Statistical methods for research workers*. London, Oliver and Boyd.
17. Neyman, J. & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society (A)*, 231, 289-337
18. Watanabe, N. & Imaizumi, T. (1993). *A fuzzy statistical test of fuzzy hypotheses*. *Fuzzy Sets and Systems*, 53, 167-178.
19. Arnold, B. F. (1995). Statistical tests optimally meeting certain fuzzy requirements on the power function and on the sample size, *Fuzzy Sets and Systems*, 75, 365-372.
20. Arnold, B. F. (1996). An approach to fuzzy hypothesis testing. *Metrika*, 44, 119-126.
21. Arnold, B. F. (1998). Testing fuzzy hypotheses with crisp data. *Fuzzy Sets and Systems*, 94, 323-333.
22. Taheri, S. M. & Behboodian, J. (1999). Neyman-Pearson lemma for fuzzy hypotheses testing, *Metrika*, 49, 3-17.
23. Arnold, B. F. & Gerke, O. (2003). Testing fuzzy linear hypotheses in linear regression models. *Metrika*, 57, 81-95.
24. Box, G. E. & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, Mass.
25. Delgado, M., Verdegay, J. L. & Vila, M. A. (1985). *Testing fuzzy hypotheses, a Bayesian approach*. In M. M. Gupta, editor, *Approximate reasoning in expert systems*, North-Holland, Amsterdam.
26. Taheri, S. M. & Behboodian, J. (2001). A Bayesian approach to fuzzy hypotheses testing. *Fuzzy Sets and Systems*, 123, 39-48.
27. Casals, M. R., Gil, M. A. & Gil, P. (1986). On the use of Zadeh's probabilistic definition for testing statistical hypotheses from fuzzy information. *Fuzzy Sets and Systems*, 20, 175-190.

28. Casals, M. R. & Gil, M. A. (1989). A note on the operativeness of Neyman-Pearson tests with fuzzy information, *Fuzzy Sets and Systems*, 30, 215-220.
29. Gil, M. A. & Casals, M. R. (1988). An operative extension of the likelihood ratio test from fuzzy data. *Statist. Hefte*, 29, 191-203.
30. Bandemer, H. & Näther, W. (1992). *Fuzzy data analysis*. Dordrecht, Kluwer Academic Publishers.
31. Filzmoser, P. & Viertl, R. (2003). Testing hypotheses with fuzzy data: the fuzzy  $p$ -value. *Metrika*, to appear.
32. Saade, J. J. & Schwarzlander, H. (1990). Fuzzy hypothesis testing with hybrid data. *Fuzzy Sets and Systems*, 35, 197-212.
33. Saade, J. J. (1994), Extension of fuzzy hypothesis testing with hybrid data. *Fuzzy Sets and Systems*, 63, 57-71.
34. Kang, M. K., Jang, S. W. & Kwon, O. J. (2001). Testing statistical hypotheses for fuzzy data with agreement index. *Far East J. Theor. Stat.* 5, 343-352.
35. Grzegorzewski, P. (2002). Testing fuzzy hypotheses with vague data. Statistical modeling, analysis and management of fuzzy data, *Stud. Fuzziness Soft Comput., Physica, Heidelberg*, 87, 213-225.
36. Casals, M. R., Gil, M. A. & Gil, P. (1986). The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information. *Eur. J. Oper. Res.* 27, 371-382.
37. Frühwirth-Schnatter, S. (1993). *On fuzzy Bayesian inference*. *Fuzzy Sets and Systems*, 60, 41-58.
38. Casals, M. R. & Gil, P. (1990). Bayes testing of fuzzy hypotheses based on fuzzy information. (*Spanish*) *Rev. Real Acad. Cienc. Exact. Fis. Natur. Madrid*, 84, 441-451.
39. Casals, M. R. (1993). Bayesian testing of fuzzy parametric hypotheses from fuzzy information. *RAIRO, Operations Research*, 27, 189-199.
40. Taheri, S. M. & Behboodan, J. (2002). *Fuzzy hypotheses testing with fuzzy data: a Bayesian approach*. In R. P. Nikhil and M. Sugeno, editors, *Advances in Soft Computing, AFSS 2002*, Berlin, Springer.

## APPENDIX

The random vector  $\mathbf{X} = (X_{1,L}, X_n)'$  is said to have an exchangeable normal distribution if its distribution is multivariate normal with the following mean vector and variance-covariance matrix [14].

$$\begin{pmatrix} \mu \\ \mu \\ \cdot \\ \mu \end{pmatrix}_{n \times 1}, \quad -\infty < \mu < \infty, \quad \sigma^2 \begin{pmatrix} 1 & \rho & L & \rho \\ \rho & 1 & L & \rho \\ \cdot & \cdot & L & \cdot \\ \rho & \rho & L & 1 \end{pmatrix}_{n \times n}, \quad \sigma > 0, \quad \rho \in [0,1),$$

We denote this exchangeable normal distribution with three parameters  $\mu, \sigma^2$ , and  $\rho$  by  $EN_n(\mu, \sigma^2, \rho)$ . It is clear that  $(X_{1,L}, X_n)$  and  $(X_{i_1,L}, X_{i_n})$  are identical in distribution for any permutation  $\{i_{1,L}, i_n\}$  of  $\{1,L, n\}$ . In the following theorem we prove that the test  $\phi_\mu(\cdot)$  in Example 3.2 is the best. (The proof for  $\phi_3$  and  $\phi_4$  is not difficult.)

**Theorem:** If  $\mathbf{X} = (X_{1,L}, X_n)'$  has the distribution  $EN_n(\mu, \sigma^2, \rho)$ , then the test  $\phi_\mu(\cdot)$  is an UMPUT for (3.4) and if  $\mu$  is unknown then there is no UMPUT for (3.4).

**Proof:** Without loss of generality assume that  $\mu = 0$ . First, consider the case  $n = 2$ . In this case the joint density of  $\mathbf{X} = (X_1, X_2)'$  is given by

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= (2\pi\sigma^2\sqrt{1-\rho^2})^{-1} \exp\left\{\frac{-1}{2(1-\rho^2)\sigma^2}(x_1^2+x_2^2-2\rho x_1x_2)\right\} \\ &= (2\pi\sigma^2\sqrt{1-\rho^2})^{-1} \exp\left\{\frac{\rho}{(1-\rho^2)\sigma^2}x_1x_2 - \frac{1}{2(1-\rho^2)\sigma^2}(x_1^2+x_2^2)\right\} \\ &= k(\theta_1, \theta_2) \exp\{\theta_1 t_1 + \theta_2 t_2\}, \end{aligned}$$

where  $\theta_1 = \frac{\rho}{(1-\rho^2)\sigma^2}$ ,  $t_1 = x_1x_2$ ,  $\theta_2 = \frac{-1}{2(1-\rho^2)\sigma^2}$ ,  $t_2 = x_1^2+x_2^2$ , and  $k(\theta_1, \theta_2)$  is a function of  $\theta_1, \theta_2$ .

Now, we can apply Theorem 3 [12]. The test function  $\phi(t_1, t_2)$  given by  $\phi(t_1, t_2) = \begin{cases} 1 & t_1 > c(t_2) \\ 0 & t_1 < c(t_2) \end{cases}$

is an UMPUT for testing  $\begin{cases} H_0^* : \theta_1 = 0 \\ H_1^* : \theta_1 > 0 \end{cases}$ , where  $c(t_2)$  is chosen such that  $P_{\theta_1=0}(T_1 > c(T_2) | T_2 = t_2) = \alpha$ . But  $\begin{cases} H_0^* : \theta_1 = 0 \\ H_1^* : \theta_1 > 0 \end{cases}$  is equivalent to  $\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$ , and on the boundary of  $H_0^*, H_1^*$  (or  $H_0, H_1$ ) i.e.  $\theta_1 = 0$  (or  $\rho = 0$ ),  $T_2$  is a complete sufficient statistic for  $\sigma^2$ . If we

define  $T'$  by

$$T' = \frac{2T_1 + T_2}{T_2} = \frac{2X_1X_2 + X_1^2 + X_2^2}{X_1^2 + X_2^2} = \frac{(X_1 + X_2)^2}{X_1^2 + X_2^2} = \frac{(\frac{X_1}{\sigma} + \frac{X_2}{\sigma})^2}{\frac{X_1^2 + X_2^2}{\sigma^2}},$$

then it is an ancillary, and as a result independent from  $T_2$ . Therefore,

$$\begin{aligned} P_{\theta_1=0}(T_1 > c(T_2) | T_2 = t_2) &= P_{\theta_1=0}(T' > c_1(T_2) | T_2 = t_2) \\ &= P_{\theta_1=0}(T' > c_2), \end{aligned}$$

where  $c_1(T_2) = \frac{2c(T_2) + T_2}{T_2}$  and  $c_2 = c_1(t_2)$  is a constant to be determined for given  $\alpha$ . Hence

$$\phi(t_1, t_2) = \begin{cases} 1 & \frac{2t_1 + t_2}{t_2} > c_2, \\ 0 & \frac{2t_1 + t_2}{t_2} < c_2, \end{cases}$$

where  $c_2$  may be chosen so that  $P_{\theta_1=0}(T' > c_2) = \alpha$ . This test is, in fact, the usual  $t$ -test, more often written in the form

$$\phi(\mathbf{x}) = \begin{cases} 1 & \frac{|\bar{x}|}{s/\sqrt{n}} > c', \\ 0 & \frac{|\bar{x}|}{s/\sqrt{n}} < c', \end{cases}$$

where  $n = 2$ ,  $|\bar{x}| = \frac{\sqrt{2t_1+t_2}}{n}$ ,  $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1) = (t_2 - n\bar{x}^2) / (n-1)$  [15]. With  $c' = t(n-1; 1-\frac{\alpha}{2})$ , the test function  $\phi(\mathbf{x})$  is an UMPU size- $\alpha$  test for testing  $\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$ .

When  $n > 2$ , the proof is similar to the above proof. In this case by simple algebraic calculations or by using an orthogonal transformation, we can show that the density of  $\mathbf{X} = (X_1, \dots, X_n)'$  is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= k_n(\rho, \sigma^2) \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^n x_i^2}{\sigma^2(1-\rho)} - \frac{\rho(\sum_{i=1}^n x_i)^2}{\sigma^2(1+(n-1)\rho)(1-\rho)}\right]\right\} \\ &= k_n(\rho, \sigma^2) \exp\left\{\frac{n\rho(\sqrt{n}\bar{x})^2}{2\sigma^2(1+(n-1)\rho)(1-\rho)} - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2(1-\rho)}\right\}, \end{aligned}$$

where  $k_n(\rho, \sigma^2) = (\sqrt{2\pi}\sigma)^{-n} (1-\rho)^{-(n-1)/2} (1+(n-1)\rho)^{-1/2}$ .

Note that,  $(\sqrt{n}\bar{x})^2 = (2\sum_{i<j} x_i x_j + \sum_{i=1}^n x_i^2) / n$ . Therefore,

$$f_{\mathbf{X}}(\mathbf{x}) = k'_n(\theta_1, \theta_2) \exp\{\theta_1 t_1 + \theta_2 t_2\},$$

where  $\theta_1 = \rho/\sigma^2(1+(n-1)\rho)(1-\rho)$ ,  $t_1 = \sum_{i<j} x_i x_j$ ,  $t_2 = \sum_{i=1}^n x_i^2$ ,  $k'_n(\theta_1, \theta_2)$  is a function of  $\theta_1$ ,  $\theta_2$ , and  $\theta_2$  can be determined. The rest of this case is similar to the case  $n = 2$ . Therefore,  $\phi_{\mu}$  is an UMPUT for (3.4).

If  $\mu$  and  $\sigma^2$  are both unknown we have a trivial UMPUT for  $\rho$ . To prove this fact we observe that

$$f_{\mathbf{X}}(\mathbf{x}) = q_n(\theta_1, \theta_2, \theta_3) \exp\{\theta_1 t_1 + \theta_2 t_2 + \theta_3 t_3\},$$

where  $\theta_1 = \rho/l$ ,  $t_1 = \sum_{i<j} x_i x_j$ ,  $\theta_2 = -(1+(n-2)\rho)/(2l)$ ,  $t_2 = \sum_{i=1}^n x_i^2$ ,  $\theta_3 = (1-\rho)\mu/l$ ,  $t_3 = \sum_{i=1}^n x_i$ ,  $l = \sigma^2(1+(n-1)\rho)(1-\rho)$ , and  $q_n(\theta_1, \theta_2, \theta_3)$  can be determined. Therefore the test function  $\phi(t_1, t_2)$  (in the first part of the proof) given by  $\phi(t_1, t_2, t_3) = \begin{cases} 1 & t_1 > c(t_2, t_3) \\ 0 & t_1 < c(t_2, t_3) \end{cases}$  is UMPU test for testing  $\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$ , where  $c(t_2, t_3)$  is chosen such that

$$P_{\theta_1=0}(T_1 > c(T_2, T_3) | T_2 = t_2, T_3 = t_3) = \alpha.$$

But note that  $T_1 = (T_3^2 - T_2)/2$ , and so the event  $\{T_1 > c(T_2, T_3)\}$  depends on  $T_2$  and  $T_3$ . Therefore, we have

$$P_{\theta_1=0}(T_1 > c(T_2, T_3) | T_2 = t_2, T_3 = t_3) = \begin{cases} 1 & t_1 > c(t_2, t_3) \\ 0 & t_1 < c(t_2, t_3) \end{cases},$$

which is equal to  $\phi(t_1, t_2, t_3)$ . (Note that if we use the method in the first part of the proof, then we obtain a similar result.)

Table 1. Trends on hypothesis testing for fixed sample size

	Crisp Observations	Fuzzy Observations
Crisp Hypotheses in Classical Framework	... Fisher (1925) [16] Neyman & Pearson (1933) [17] ...	Casals et al. (1986,1989) [27, 28] Gil & Casals (1988) [29] Bandemer & Näther (1992) [30] Filzmoser & Viertl (2003) [31]
Fuzzy Hypotheses in Classical Framework	Watanabe & Imaizumi (1993) [18] Arnold (1995,1996,1998) [19, 20, 21] Taheri & Behboodian (1999) [22] Arnold & Gerke (2003) [23]	Saade & Schwarzlander (1990) [32] Saade (1994) [33] Kang et al. (2001) [34] Grzegorzewski (2002) [35]
Crisp Hypotheses in Bayesian Framework	... Box & Tiao (1973) [24] Cox & Hinkley (1974) [11] ...	Casals et al. (1986) [36] Frühwirth-Schnatter (1993) [37]
Fuzzy Hypotheses in Bayesian Framework	Delgado et al. (1985) [25] Taheri & Behboodian (2001) [26]	Casals & Gil (1990) [38] Casals (1993) [39] Taheri & Behboodian (2002) [40]