Testing the expected value of a fuzzy random variable. A discussion

Manuel Montenegro
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo, 33071 Oviedo, Spain
manuel@pinon.ccu.uniovi.es

Mª Rosa Casals
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo, 33071 Oviedo, Spain
rosa@pinon.ccu.uniovi.es

Ana Colubi
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo, 33071 Oviedo, Spain
colubi@pinon.ccu.uniovi.es

Mª Ángeles Gil
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo, 33071 Oviedo, Spain
angeles@pinon.ccu.uniovi.es

Abstract

In this communication we will consider hypothesis-tests for the (fuzzy-valued) mean value of a fuzzy random variable in a population. For this purpose, we will make use of a generalized metric for fuzzy numbers, and we will develop two different approaches for the case of fuzzy random variables taking on a finite number of possible values, both leading to close statistical inferences.

Keywords: bootstrap technique; fuzzy random variable; mean value of a fuzzy random variable; large sample theory.

1 Introduction

In this communication we consider the problem of testing the fuzzy value of the mean of a fuzzy-valued random variable taking on values on the class $\mathcal{F}_c(\mathbb{R})$. This class denotes the collection of fuzzy subsets $\widetilde{A}$ of $\mathbb{R}$ such that the α-level sets $\widetilde{A}_\alpha \in \mathcal{K}_c(\mathbb{R})$ ($\mathcal{K}_c(\mathbb{R})$ being the class of nonempty compact intervals) for all $\alpha \in [0,1]$, with the 0-level meaning the closure of the support of $\widetilde{A}$.

The first study we will present concerns large-sample tests for simple fuzzy random variables. Therefore, there are three prerequisites to using these tests, namely, that the samples are assumed to be large and random, and that there is only a finite number of different categories for the considered fuzzy random variable. The last prerequisite is not general from a theoretical perspective, but it is very realistic and fits most of the practical situations involving fuzzy random variables.

In the second study, by using bootstrap techniques, we obtain an approximation of the asymptotic distribution of a statistic based on a resampling from the reference sample. Simulation developements will allow us to conclude that the approximation based on the asymptotic results is usually as appropriate as the bootstrap one even for non-large samples.

2 Preliminaries

Given a probability space $(\Omega, \mathcal{A}, P)$, a *fuzzy random variable* (or, more precisely, an $\mathcal{F}_c(\mathbb{R})$-valued fuzzy random variable or random fuzzy set) associated with $(\Omega, \mathcal{A})$ is intended to be, in accordance with Puri and Ralescu (1986), a function $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$ such that the α-level mapping $\mathcal{X}_\alpha : \Omega \to \mathcal{K}_c(\mathbb{R})$, defined so that $\mathcal{X}_\alpha(\omega) = (\mathcal{X}(\omega))_\alpha$ for all $\omega \in \Omega$, is a random compact convex set whatever $\alpha \in [0,1]$ may be.

A fuzzy random variable $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$ is said to be *integrably bounded* if, and only if, $\|\mathcal{X}_0\| \in L^1(\Omega, \mathcal{A}, P)$ (with $\|\mathcal{X}_0\|(\cdot) = \sup_{x \in \mathcal{X}_0(\cdot)} |x|$). If $\mathcal{X}$ is an integrably bounded fuzzy random variable, the *expected value* (or fuzzy mean) of $\mathcal{X}$ is the unique fuzzy set of $\mathbb{R}$, $\widetilde{E}(\mathcal{X}|P)$, such that $(\widetilde{E}(\mathcal{X}|P))_\alpha$ = Aumann's integral (1965) of $\mathcal{X}_\alpha$ (which in this case equals the compact interval $[E(\inf \mathcal{X}_\alpha|P), E(\sup \mathcal{X}_\alpha|P)]$) for all $\alpha \in [0,1]$.

A fuzzy random variable $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$ is said to be *normal* in Puri and Ralescu's sense (1986) if, and only if, for each $\alpha \in [0,1]$ and each $\omega \in \Omega$ the compact set $\mathcal{X}_\alpha(\omega)$ can be expressed as the Minkowski sum of the real value $X(\omega)$ and the α-level of the fuzzy set $\widetilde{E}(\mathcal{X}|P) \in \mathcal{F}_c(\mathbb{R})$ (that is, $\mathcal{X}(\cdot) = X(\cdot) \oplus \widetilde{E}(\mathcal{X}|P)$ with $\oplus$ the fuzzy sum according to Zadeh's extension principle [4]), where $X : \Omega \to I\!R$ is a real-valued random variable associated with $(\Omega, \mathcal{A}, P)$ and being normally distributed with mean equal to 0.

A fuzzy random variable $\mathcal{X} : \Omega \to \mathcal{F}_c(I\!R)$ is said to be *simple* if, and only if, the cardinality of $\mathcal{X}(\Omega)$ is finite.

On the class $\mathcal{F}_c(I\!R)$ several metrics can be defined (see, for instance, Diamond and Kloeden 1994). For some statistical studies concerning fuzzy random variables, a generalized metric has been shown to be especially valuable and easy to handle and interpret. This metric has been introduced by Bertoluzza *et al.* (1995) and a generalization has been recently established by Körner and Näther (2001). The (W, φ)-metric by Bertoluzza *et al.* makes use of two previously specified weight normalized measures W and φ which can be formalized by means of probability measures on the measurable space $([0,1], \mathcal{B}_{[0,1]})$ ($\mathcal{B}_{[0,1]}$ being the Borel σ-field on $[0,1]$) associated with a nondegenerate and a continuous distribution, respectively, and the (W, φ)-distance between two fuzzy numbers $\widetilde{A}, \widetilde{B} \in \mathcal{F}_c(I\!R)$ is defined by

$$D_W^{\varphi}(\widetilde{A}, \widetilde{B}) = \sqrt{\int_{[0,1]} [d_W(\widetilde{A}_\alpha, \widetilde{B}_\alpha)]^2 d\varphi(\alpha)},$$

where $d_W : \mathcal{K}_c(I\!R) \times \mathcal{K}_c(I\!R) \to [0, +\infty)$ is defined so that $d_W(\widetilde{A}_\alpha, \widetilde{B}_\alpha)$ is given by

$$\sqrt{\int_{[0,1]} \left[f_{\widetilde{A}}(\alpha, \lambda) - f_{\widetilde{B}}(\alpha, \lambda)\right]^2 dW(\lambda)}$$

with $f_{\widetilde{A}}(\alpha, \lambda) = \lambda \sup \widetilde{A}_\alpha + (1 - \lambda) \inf \widetilde{A}_\alpha$.

3 Asymptotic one-sample test of mean for simple fuzzy random variables

Classical inferences to test the population mean of a real-valued random variable are often based (but in cases normality is assumed) on the asymptotic distribution of a statistic which is obtained under quite general conditions by using Large Sample Theory (more precisely, properties of the sample means). When we deal with fuzzy random variables, the situation becomes much more complex since, in contrast to the case of real-valued random variables, population and sample means are not real-valued.

However, the asymptotic distribution is going to be obtained in this section under conditions which are very common in practice. Thus, if we assume that the considered fuzzy random variable we are observing in the population is a simple one, we have a vector-valued parameter for which results from Large Sample Theory can be directly applied. On the other hand, the assumption that the fuzzy random variable is simple is quite realistic in practice, so that the asymptotic results we now present will be widely applicable.

Consider a probability space $(\Omega, \mathcal{A}, P)$. Let $\mathcal{X}$ be a fuzzy random variable associated with it, so that on Ω the fuzzy random variable takes on r different values, $\widetilde{x}_1, \ldots, \widetilde{x}_r$ with probabilities $p_1, \ldots, p_r$, respectively (i.e., $P(\{\omega \in \Omega | \mathcal{X}(\omega) = \widetilde{x}_l\} = p_l$, $l = 1, \ldots, r$).

The population fuzzy mean of $\mathcal{X}$ over Ω is now denoted by $\widetilde{E}(\mathcal{X}|\mathbf{p})$ (where $\mathbf{p}$ is the vector-valued parameter $\mathbf{p} = (p_1, \ldots, p_{r-1})$ and $p_r = 1 - \sum_{l=1}^{r-1} p_l$). Then,

Theorem 3.1 *For each $n \in I\!N$, consider n independent fuzzy random variables having identical distribution that $\mathcal{X}$ on Ω.*

Let $\mathbf{f}_n = (f_{n1}, \ldots, f_{n(r-1)})$ with f_{nl} = relative frequency of $\widetilde{x}_l$ ($l \in \{1, \ldots, r-1\}$) in the performance of the n fuzzy random variables. Let $\overline{\mathcal{X}} = \widetilde{E}(\mathcal{X}|\mathbf{f}_n)$ denote the sample fuzzy mean of the n fuzzy random variables. Then, to test at the significance level $\alpha \in [0,1]$ the null hypothesis

$$H_0 : \widetilde{E}(\mathcal{X}|\mathbf{p}) = \widetilde{V}$$

against the alternative

$$H_a : \widetilde{E}(\mathcal{X}|\mathbf{p}) \neq \widetilde{V},$$

the hypothesis H_0 should be (asymptotically) rejected whenever

$$2n\left[D_W^{\varphi}(\overline{\mathcal{X}}, \widetilde{V})\right]^2 > \gamma_\alpha,$$

where γ_α is the $100(1-\alpha)$ fractile of the linear combination of chi-square independent variables $\widehat{\lambda}_1 \chi^2_{1,1} + \ldots + \widehat{\lambda}_k \chi^2_{1,k}$, with $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k$ ($k \leq r-1$) being the nonnull (sample) eigenvalues of the matrix

$$B^t H\left(\left[D_W^{\varphi}(\overline{\mathcal{X}}, \widetilde{V})\right]^2\right) B,$$

where $H\left(\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^{2}\right)$ *is the Hessian matrix*

$$H\left(\left[D_{W}^{\varphi}(\widetilde{E}(\mathcal{X}|\mathbf{f}_n),\widetilde{V})\right]^{2}\right)$$

$$= \begin{pmatrix} \frac{\partial^2\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^2}{\partial f_{n1}\partial f_{n1}} & \cdots & \frac{\partial^2\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^2}{\partial f_{n1}\partial f_{n(r-1)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^2}{\partial f_{n(r-1)}\partial f_{n1}} & \cdots & \frac{\partial^2\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^2}{\partial f_{n(r-1)}\partial f_{n(r-1)}} \end{pmatrix},$$

and B *is an* $(r-1)\times(r-1)$ *matrix such that* $B^tB = \left(I_{\mathcal{X}}^{F}(\mathrm{f}_n)\right)^{-1}$, *where* $(I_{\mathcal{X}}^{F}(\mathrm{f}_n))^{-1}$ *is the inverse of the sample Fisher information matrix* $[f_{nl}(\delta_{lm}-f_{nm})]_{lm}$.

4 Bootstrap one-sample test of mean for simple fuzzy random variables

An alternative way to get suitable approximations in statistical inferences with real-valued random variables is the one based on the bootstrap technique. When we deal with simple fuzzy random variables and consider a real-valued statistic based on samples from a fuzzy random variable, we can develop this technique as follows:

Consider a probability space $(\Omega,\mathcal{A},P)$. Let $\mathcal{X}$ be a fuzzy random variable associated with it, so that on Ω the fuzzy random variable takes on r different values, $\tilde{x}_1,\ldots,\tilde{x}_r$ with probabilities $p_1,\ldots,p_r$, respectively. For each $n\in\mathbb{N}$, consider a random sample of n independent observations from $\mathcal{X}$, and let $\mathrm{f}_n=(f_{n1},\ldots,f_{nr})$ be the random vector of the sample frequencies.

Consider now a new population given by a realization of the preceding random sample. Assume that we resample from it so that we draw a large number of samples of n independent observations from this new population. If we denote the associated random vector of the sample frequencies along the set of these samples by $\mathbf{f}_n^*=(f_{n1}^*,\ldots,f_{nr}^*)$, we can then obtain the bootstrap distribution which in this case corresponds to the approximation of the one for $\left[D_{W}^{\varphi}(\widetilde{E}(\mathcal{X}|\mathrm{f}_n),\widetilde{E}(\mathcal{X}|\mathbf{f}_n^*))\right]^2$ by MonteCarlo method.

Theorem 4.1 *To test at the significance level* $\alpha\in[0,1]$ *the null hypothesis*

$$H_0:\widetilde{E}(\mathcal{X}|\mathbf{p})=\widetilde{V}$$

against the alternative

$$H_{\mathrm{a}}:\widetilde{E}(\mathcal{X}|\mathbf{p})\neq\widetilde{V},$$

the hypothesis H_0 *should be (in accordance with the bootstrap approximation) rejected whenever*

$$\left[D_{W}^{\varphi}(\overline{\mathcal{X}},\widetilde{V})\right]^2 > z_\alpha,$$

where z_α *is the* $100(1-\alpha)$ *fractile of the bootstrap distribution.*

5 Simulation conclusions

To compare the results in the two preceding sections, we have developed a study in which some different types of fuzzy random variables have been simulated by using the ideas in Colubi *et al.* (2001). The main conclusions from the simulation studies can be summarized as follows:

- If $r=100$, $\mathbf{p}=(.01,\ldots,.01)$, and significance level $\alpha=.05$, then,
  - for a sample size $n=50$, we obtain that when the null hypothesis is true it is rejected for 7% of the samples if we use either the approach in Section 3 or the one in Section 4, whereas if we were able to know the population eigenvalues in Theorem 3.1 this rejection will reduce to 6%;
  - for a sample size $n=10$, we obtain that when the null hypothesis is true it is rejected for 10% of the samples if we use any of the approaches in Sections 3 and 4, whereas if we know the population eigenvalues this rejection reduces to 4.5%.
- If $r=5$, $\mathbf{p}=(.02,.02,.02,.02)$, and significance level $\alpha=.05$, then,
  - for a sample size $n=50$, we obtain that when the null hypothesis is true it is rejected for 6.5% of the samples if we use the approach in Section 3, for 7% if we consider the one in Section 4, and for 5.5% in case we know the population eigenvalues;

- for a sample size $n = 10$, we obtain that when the null hypothesis is true it is rejected for 9.5% of the samples if we use the approach in Section 3, for 10% if we consider the one in Section 4, and for 5% in case we know the population eigenvalues.

- If $r = 5$, components of $\mathbf{p}$ are different, and significance level $\alpha = .05$, then,
  - for a sample size $n = 50$, we obtain that when the null hypothesis is true it is rejected for 7% of the samples if we use the approach in Section 3, for 6.5% if we consider the one in Section 4, and for 6% in case we know the population eigenvalues;
  - for a sample size $n = 10$, we obtain that when the null hypothesis is true it is rejected for 8.5% of the samples if we use the approach in Section 3, for 9% if we consider the one in Section 4, and for 4% in case we know the population eigenvalues.

In the simulations above summarized, the measures W and φ have been supposed to be the Lebesgue measure on the unit interval.

Acknowledgements

This work has been partially supported by two Grants from the Spanish Ministerio de Educación, Cultura y Deporte (DGE-99-PB98-1534 and DGE-98-PB97-1282). Their financial support is gratefully acknowledged. The authors wish also to thank their colleague Gil González for his scientifical support and suggestions on the results in Sections 4 and 5.

References

[1] R.J. Aumann (1965), Integrals of set-valued functions, *J. Math. Anal. Appl.* **12**, 1–12.

[2] C. Bertoluzza, N., Corral and A. Salas (1995), On a new class of distances between fuzzy numbers, *Mathware & Soft Computing* **2**, 71–84.

[3] A. Colubi, C. Fernández-García and M.A. Gil (2001), Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data, (submitted for publication).

[4] P. Diamond and P. Kloeden (1994), *Metric Spaces of Fuzzy Sets: Theory and Applications*, World Scientific, Singapore.

[5] R. Körner and W. Näther (2001), On the variance of random fuzzy variables, in *Statistical Modeling, Analysis and Management of Fuzzy Data* (C. Bertoluzza, M.A. Gil and D.A. Ralescu, Eds.), Physica-Verlag, 22-39.

[6] M.L. Puri and D.A. Ralescu (1986), Fuzzy random variables, *J. Math. Anal. Appl.* **114**, 409–422.