A Method to Simulate Fuzzy Random Variables

González-Rodríguez. G.[1], Colubi, A.[1], Gil, M.A.[1] and Coppi, R.[2]

[1] Dpto. de Estadística e I.O. Universidad de Oviedo. 33007 Spain
{gil,colubi,magil}@uniovi.es
[2] Dpto. di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma, "La Sapienza"
renato.coppi@uniroma1.it

In this paper a method is introduced to simulate fuzzy random variables by using the support function. On the basis of the support function, the class of values of a fuzzy random variable can be 'identified' with a closed convex cone of a Hilbert space, and we now suggest to simulate Hilbert space-valued random elements and to project later into such a cone. To make easier the projection above we will consider isotonic regression. The procedure will be illustrated by means of several examples.

1 Introduction

In the literature on fuzzy-valued random variables, there are only a few references to modeling the distribution of these random elements. These models (for instance, see [12]) are theoretically well stated, but they are not soundly supported by empirical evidence, since they correspond to quite restrictive random mechanisms and hence they are not realistic in practice (see [4]).

Nevertheless, many probabilistic and statistical studies on fuzzy random variables would be better developed if simulation studies could be carried out (cf. [8], [9], [11]).

A similar situation arises in connection with functional data, to which a lot of attention is being paid in the last years, especially in which concerns random elements taking on values in Hilbert spaces (see, for instance, [14], [15]). The assumption of the Hilbert space structure is very helpful for simulation purposes (see [7], [1] or [16]).

The key idea in the methodology to be presented is first based on passing from the space of fuzzy random variable values into the Hilbert space of the corresponding integrable functions through the support function; then, one can generate Hilbert space-valued random elements and project them into the convex cone of the image of the space of fuzzy values. The projection theorem in Hilbert spaces validates the way to proceed and theoretically it would be possible to simulate all possible distributions on the space.

G. González-Rodríguez et al.: *A Method to Simulate Fuzzy Random Variables*, Advances in Soft Computing **6**, 103–110 (2006)
www.springerlink.com

This idea can be easily implemented from a theoretical viewpoint. In practice, when fuzzy values to be dealt with are fuzzy sets of the one-dimensional Euclidean space the implementation does not entail important difficulties, since the support function of a fuzzy value is characterized by two real-valued functions on the unit interval, namely, the one associated with the infima and that associated with the suprema. These two functions are in the cone of the monotonic functions, and they are subject to the constraint of the infimum being lower than the supremum for each level. They have been analyzed in connection with some probabilistic problems (see [2]). However, for fuzzy sets of multi-dimensional Euclidean spaces, the practical developments become much more complex, although some alternatives to simplify them will be commented along the paper.

In this paper a procedure to simulate fuzzy random variables for which the shape of fuzzy values is not constrained will be introduced. In case there are some preferences on the shape of the considered fuzzy values the procedure could also adapted.

2 Preliminaries

Let $\mathscr{K}_c(\mathbb{R}^p)$ be the class of the nonempty compact convex subsets of $\mathbb{R}^p$ endowed with the Minkowski sum and the product by a scalar, that is, $A+B=\{a+b\,|\,a\in A,\, b\in B\}$ and $\lambda A=\{\lambda a\,|\,a\in A\}$ for all $A,B\in\mathscr{K}_c(\mathbb{R}^p)$ and $\lambda\in\mathbb{R}$. We will consider the *class of fuzzy sets*

$$\mathscr{F}_c(\mathbb{R}^p)=\left\{U:\mathbb{R}^p\to[0,1]\;\middle|\;U_\alpha\in\mathscr{K}_c(\mathbb{R}^p)\text{ for all }\alpha\in[0,1]\right\}$$

where U_α is the α-level of U (i.e. $U_\alpha=\{x\in\mathbb{R}^p\,|\,U(x)\geq\alpha\}$) for all $\alpha\in(0,1]$, and U_0 is the closure of the support of U. The space $\mathscr{F}_c(\mathbb{R}^p)$ can be endowed with the sum and the product by a scalar based on Zadeh's extension principle [17], which satisfies that $(U+V)_\alpha=U_\alpha+V_\alpha$ and $(\lambda U)_\alpha=\lambda U_\alpha$ for all $U,V\in\mathscr{F}_c(\mathbb{R}^p)$, $\lambda\in\mathbb{R}$ and $\alpha\in[0,1]$.

The support function of a fuzzy set $U\in\mathscr{F}_c(\mathbb{R}^p)$ is $s_U(u,\alpha)=\sup_{w\in U_\alpha}\langle u,w\rangle$ for any $u\in\mathbb{S}^{p-1}$ and $\alpha\in[0,1]$, where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$ and $\langle\cdot,\cdot\rangle$ denotes the inner product. The support function allows us to embed $\mathscr{F}_c(\mathbb{R}^p)$ onto a cone of the continuous and Lebesgue integrable functions $\mathscr{L}(\mathbb{S}^{p-1})$ by means of the mapping $s:\mathscr{F}_c(\mathbb{R}^p)\to\mathscr{L}(\mathbb{S}^{p-1}\times[0,1])$ where $s(U)=s_U$ (see [5]).

We will consider the *generalized metric* by Körner and Näther [10] D_K, which is defined so that

$$[D_K(U,V)]^2=$$

$$\int_{(\mathbb{S}^{p-1})^2\times[0,1]^2}\big(s_U(u,\alpha)-s_V(u,\alpha)\big)\big(s_U(v,\beta)-s_V(v,\beta)\big)\,dK(u,\alpha,v,\beta),$$

for all $U,V\in\mathscr{F}_c(\mathbb{R}^p)$, where K is a positive definite and symmetric kernel; thus, D_K coincides with a generic L_2 distance $\|\cdot\|_2$ on the Hilbert space $\mathscr{L}(\mathbb{S}^{p-1}\times[0,1])$.

Let $(\Omega,\mathscr{A},P)$ be a probability space. A *fuzzy random variable* (FRV) in Puri & Ralescu's sense [13] is a mapping $\mathscr{X}:\Omega\to\mathscr{F}_c(\mathbb{R}^p)$ so that the α-level mappings

$\mathscr{X}_\alpha : \Omega \to \mathscr{K}_c(\mathbb{R}^p)$, defined so that $\mathscr{X}_\alpha(\omega) = \big(\mathscr{X}(\omega)\big)_\alpha$ for all $\omega \in \Omega$, are random sets (that is, Borel-measurable mappings with the Borel σ-field generated by the topology associated with the well-known Hausdorff metric d_H on $\mathscr{K}(\mathbb{R}^p)$). Alternatively, an FRV is an $\mathscr{F}_c(\mathbb{R}^p)$-valued random element (i.e. a Borel-measurable mapping) when the Skorokhod metric is considered on $\mathscr{F}_c(\mathbb{R}^p)$ (see [3]).

If $\mathscr{X} : \Omega \to \mathscr{F}_c(\mathbb{R}^p)$ is a fuzzy random variable such that $d_H\big(\{0\}, \mathscr{X}_0\big) \in L^1(\Omega, \mathscr{A}, P)$, then the *expected value (or mean)* of $\mathscr{X}$ is the unique $E(\mathscr{X}) \in \mathscr{F}_c(\mathbb{R}^p)$ such that $\big(E(\mathscr{X})\big)_\alpha =$ Aumman's integral of the random set $\mathscr{X}_\alpha$ for all $\alpha \in [0,1]$, that is,

$$\big(E(\mathscr{X})\big)_\alpha = \big\{E(X|P) \;\big|\; X : \Omega \to \mathbb{R}^p, X \in L^1(\Omega, \mathscr{A}, P), X \in \mathscr{X}_\alpha \; a.s. \; [P]\big\}.$$

3 Simulation of Fuzzy Random Variables Through Functional Random Variables

The space of fuzzy values $\mathscr{F}_c(\mathbb{R}^p)$ is a closed convex cone of the Hilbert space $\mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$, and hence there exists a unique projection. As a consequence, given an arbitrary $f \in \mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$ there is a unique fuzzy set $P(f) = A_f$ which corresponds to the anti-image of the support function of the projection of f onto the cone $s(\mathscr{F}_c(\mathbb{R}^p))$. We will denote by $P : \mathscr{L}(\mathbb{S}^{p-1} \times [0,1]) \to s(\mathscr{F}_c(\mathbb{R}^p))$ the projection function.

For any random element X taking on values in $\mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$, the mapping $s^- 1 \circ P \circ X$ is a fuzzy random variable. In this way, if random elements of $\mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$ are generated, random elements of $s(\mathscr{F}_c(\mathbb{R}^p))$ could be obtained through the projection P. Due to the fact that $s(\mathscr{F}_c(\mathbb{R}^p)) \subset \mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$, we can guarantee that this method involves all the possible distributions on $s(\mathscr{F}_c(\mathbb{R}^p))$ and, since s is an isometry, by applying s^{-1} we would get all the possible distributions on $\mathscr{F}_c(\mathbb{R}^p)$.

The *theoretical method to generate $\mathscr{F}_c(\mathbb{R}^p)$-valued fuzzy random variables* consists in

Step 1

Simulating random elements on $\mathscr{L}(\mathbb{S}^{p-1} \times [0,1])$ by following the current directions in Functional Data Analysis (i.e., by considering bases either from a given function plus a noise term, or from discretized brownian motions, and so on).

Step 2

Projecting the simulated elements into the isometric cone of $\mathscr{F}_c(\mathbb{R}^p)$.

Step 3

Identifying the fuzzy set associated with the generated support function.

This theoretical method seems to be complex to implement in practice, although it would be feasible in some particular cases. Thus, in case $p = 1$, the unit sphere S^{p-1} reduces to the set $\{-1, 1\}$ whence the fuzzy set $A \in \mathscr{F}_c(\mathbb{R})$ can be characterized by means of two monotonic functions $s_A(-1, \cdot)$ and $s_A(1, \cdot)$ (see [2]) which satisfy certain constraints (since the infimum should always be lower than the supremum).

To make the problem easy to handle, fuzzy values can be reparameterized in terms of the left and right spreads with respect to the center of the 1-level. Once fuzzy values are reparameterized in such a way, arbitrary functions can be generated to construct later the function of the left spreads (for the infima) and the function of the right spreads (for the suprema). Since these two functions are monotonic and nonnegative, we can apply an algorithm of the isotonic regression restricted to positive values (see [6]). Later, the mid point of the 1-level would be generated at random and, along with the spreads simulated before, the infimum and supremum functions defining the fuzzy value would be obtained.

The *'practical' method to generate* $\mathscr{F}_c(\mathbb{R})$*-valued fuzzy random variables* we suggest in this paper can be summarized as follows:

Step $\mathscr{F}_c(\mathbb{R})$*-1*

To generate at random the mid-point of the 1-level, x_0, as well as two random functions on the Hilbert space $\mathscr{L}([0,1])$, $f_l, f_r : [0,1] \to R$ (there is no need for these functions to be generated independently).

Step $\mathscr{F}_c(\mathbb{R})$*-2*

To find the antitonic regressions of f_l^* and f_r^* to get the left and right spreads $s_l, s_r : [0,1] \to [0,\infty)$, respectively.

Step $\mathscr{F}_c(\mathbb{R})$*-3*

The α-levels of the fuzzy value A generated through *Steps* $\mathscr{F}_c(\mathbb{R})$*-1* and $\mathscr{F}_c(\mathbb{R})$*-2* would be given by $A_\alpha = [x_0 - s_l(\alpha), x_0 + s_r(\alpha)]$ (which is well-defined).

As we have commented before, the procedure above does not involve constraints on the shape of fuzzy values to be generated, although this type of constraint (like, for instance, to assume that x_0 is deterministic, functions f_i are linear functions, etc.) could be incorporated if required.

4 Some Illustrative Examples

We now illustrate the ideas in Section 3 by means of two examples. Since *Steps* $\mathscr{F}_c(\mathbb{R})$*-2* and $\mathscr{F}_c(\mathbb{R})$*-3* do not involve any random process, the differences in applying the algorithm are restricted to Step $\mathscr{F}_c(\mathbb{R})$-1. There are many ways of simulating random functions in the Hilbert space $\mathscr{L}([0,1])$. Some of them, as those based on a function plus a noise term or considering a class depending on real random parameters, can be easily imitated in $\mathscr{F}_c(\mathbb{R})$. However, the Hilbert spaces present some distinguishing characteristics, such as the generating basis, that can be taken into account to simulate random elements in a wider context.

In this section two ways of simulating from generating bases the functions f_1 and f_2 in *Step* $\mathscr{F}_c(\mathbb{R})$*-1* of the above-described procedure are detailed.

Consider a referential triangular fuzzy set $\mathrm{Tri}(-1,0,1)$, which is equivalent to consider the spread functions $f_1(\alpha) = f_2(\alpha) = 1 - \alpha$ for all $\alpha \in [0,1]$. Since these spread functions correspond to linear functions, the trigonometric basis will be suitable to represent them. This basis is given by

$$\varphi_j(x) = \begin{cases} 1 & \text{if } j = 0 \\ \sqrt{2}\cos(\pi j x) & \text{if } j = 1, 2, \ldots \end{cases}$$

Coefficients of the spread functions in this basis are given by

$$\theta_j = \begin{cases} .5 & \text{if } j = 0 \\ 0 & \text{if } j \text{ is an even number} \\ \dfrac{2\sqrt{2}}{\pi^2 j^2} & \text{if } j \text{ is an odd number} \end{cases}$$

For practical purposes we will consider the approximation of the function corresponding to the first 21 terms of the linear combination (i.e., $j = 0, \ldots, 20$). Coefficients are distorted in a random way so that all the generated random functions follow the expression

$$\sum_{j=0}^{20} (\theta_j + \varepsilon_j)\varphi_j$$

where $(\varepsilon_0, \ldots, \varepsilon_{20})$ is a random vector.

The way of distorting the coefficients is crucial, since small perturbations can produce shapes completely different from the original one. It should be recalled that, in order to get well-defined fuzzy sets, we will need to apply an antitonic regression algorithm after the simulation of the functions in $\mathscr{L}([0,1])$. Thus, if the simulated functions are highly variable (in the sense of showing many monotonicity changes), the antitonic regression corresponding to the spreads will have many constant parts, and hence the obtained fuzzy set will present a lot of discontinuities. In order to illustrate this behaviour, we will firstly consider the following:

Case A. For the left spread a sequence of independent realizations, $\varepsilon_0^l, \ldots, \varepsilon_{20}^l$, are simulated from the normal distribution $\mathscr{N}(0, .01)$, and for the right spread a sequence of independent realizations, $\varepsilon_0^r, \ldots, \varepsilon_{20}^r$, are simulated from the normal distribution $\mathscr{N}(0, .1)$. Thus, we get two random functions

$$f_l = \sum_{j=0}^{20} (\theta_j + \varepsilon_j^l)\varphi_j \text{ and } f_r = \sum_{j=0}^{20} (\theta_j + \varepsilon_j^r)\varphi_j.$$

The mid-point of the 1-level is chosen at random from a normal distribution $\mathscr{N}(2,1)$. By applying *Steps* $\mathscr{F}_c(\mathbb{R})$*-2* and $\mathscr{F}_c(\mathbb{R})$*-3*, a random fuzzy set is obtained.

In order to compare some particular realizations of the simulated fuzzy random variable with the expected value of such an element, we have made 10,000 simulations and we have approximated the (fuzzy) mean value by Monte Carlo method. In Figures 1 and 2 three simulated values and the corresponding mean value are shown. We can see that, although the perturbations were chosen to follow distributions with

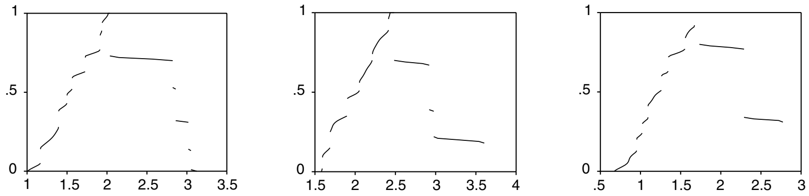

Fig. 1. Simulated values in **Case A**

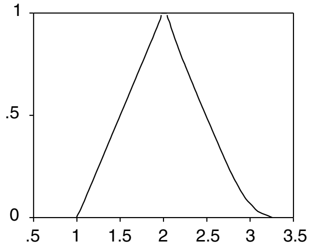

Fig. 2. Approximated mean value in the simulation in **Case A**

a relative small variability, the simulated fuzzy sets are quite different from the referential triangular fuzzy number and have many discontinuities. Nonetheless, the shape of the expected value is quite smooth and more similar to the referential fuzzy number. The difference between this mean value and the original triangular one is mainly due to the application of the antitonic regression algorithm (the expected value of the antitonic regression can be different from the antitonic regression of the expected value).

In order to obtain smoother shapes, we can simulate the perturbations in the coefficients with a decreasing weight as follows.

Case B. For the left spread a sequence of independent realizations $U_0^l, \ldots, U_{20}^l$ from the uniform distribution $\mathscr{U}_{(0,1)}$ are simulated, and the perturbations are considered so that $\varepsilon_0^l = U_0^l$, $\varepsilon_j^l = U_j^l \cdot \varepsilon_{j-1}^l$. For the right spread the same process is followed but using the beta distribution $\beta(5,3)$ instead of the uniform one. Again, the mid-point is chosen at random from a normal distribution $\mathscr{N}(2,1)$ and Steps $\mathscr{F}_c(\mathbb{R})$-2 and $\mathscr{F}_c(\mathbb{R})$-3 are followed to get the random fuzzy set. In Figures 3 and 4 three simulated values and the corresponding mean value (approximated by 10,000 realizations of the process) are shown. As expected, we can see smoother shapes than those in Case A, although they are also quite different and the greater the magnitude of right perturbations the greater the probability of discontinuities.

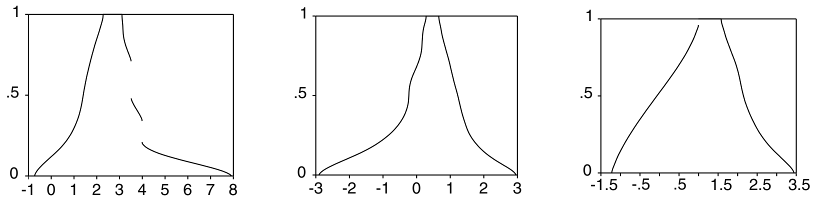

Fig. 3. Simulated values in **Case B**

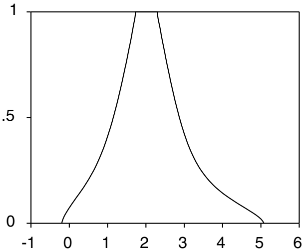

Fig. 4. Approximated mean value in the simulation in **Case B**

Acknowledgement

The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2005-00045. Its financial support is gratefully acknowledged.

References

[1] H. Cardot. Nonparametric regression for functional responses with application to conditional functional principal components analysis. Technical Report LSP-2005-01, Universite Paul Sabatier, 2005.

[2] A. Colubi, J. S. Domínguez-Menchero, M. López-Díaz, and R. Körner. A method to derive strong laws of large numbers for random upper semicontinuous functions. *Statist. Probab. Lett.*, 53:269–275, 2001.

[3] A. Colubi, J. S. Domínguez-Menchero, M. López-Díaz, and D. A. Ralescu. A $d_e[0,1]$-representation of random upper semicontinuous functions. *Proc. Amer. Math. Soc.*, 130:3237–3242, 2002.

[4] A. Colubi, C. Fernández-García, M. López-Díaz, , and M. A. Gil. Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. *IEEE Trans. Fuzzy Syst.*, 10:384–390, 2002.

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

[6] J.S. Domínguez-Menchero and G. González-Rodríguez. Analyzing an extension of the isotonic regression problem. 2006. (submitted).

[7] R. Fraiman and G. Muñiz. Trimmed means for functional data. *Test*, 10:419–440, 2001.

[8] M.A. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, and M.R. Casals. Bootstrap approach to the multi-sample test of means with imprecise data. *Comput. Statist. Data Anal.*, 2006. (accepted, in press).

[9] G. González-Rodríguez, A. Colubi, and M.A. Gil. A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. *Comput. Statist. Data Anal.*, 2006. (accepted, in press).

[10] R. Körner and W. Näther. On the variance of random fuzzy variables. In C. Bertoluzza, M.A. Gil, and D.A. Ralescu, editors, *Statistical Modeling, Analysis and Management of Fuzzy Data*, pages 22–39. Physica-Verlag, Heidelberg, 2002.

[11] M. Montenegro, A. Colubi, M. R. Casals, and M. A. Gil. Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika*, 59:31–49, 2004.

[12] M. L. Puri and D. A. Ralescu. The concept of normality for fuzzy random variables. *Ann. Probab.*, 11:1373–1379, 1985.

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

[14] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer–Verlag, New York, 1997.

[15] J. O. Ramsay and B. W. Silverman. *Applied functional data analysis. Methods and case studies*. Springer Series in Statistics. Springer–Verlag, New York, 2002.

[16] F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.*, 100:577–590, 2005.

[17] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, II. *Inform. Sci.*, 8:301–353, 1975.