ON EXPECTATION AND VARIANCE OF RANDOM FUZZY SETS, WITH APPLICATION TO LINEAR REGRESSION WITH VAGUE OBSERVATIONS

Wolfgang NÄTHER

Mining Academy of Freiberg
Department of Mathematics
9200 Freiberg, GDR

1. Introduction

Let us start with a simple (non-fuzzy) regression example. Assume, in a certain situation a linear regression model

$$Y(x) = f(x)^T\vartheta + \varepsilon \quad ; \; E\varepsilon = 0 \; ; \; \mathrm{Var}\,\varepsilon = \sigma^2; \; x \in R^k; \vartheta \in R^r \tag{1}$$

is considered and we are interested in estimation of a linear functional of ϑ, say

$$\eta = c^T\vartheta \quad ; \quad c \in R^r . \tag{2}$$

Then, based on n (point-shaped) observations $Y(x_i) = Y_i$, in the most cases the well-known least squares estimator (LSE)

$$\check{\eta} = \sum_{i=1}^{n} \check{\lambda}_i Y_i = \check{\lambda}^T Y \tag{3}$$

$$\check{\lambda}^T = c^T(F^TF)^{-1}F^T \; ; \; F = (f(x_1),\ldots,f(x_n))^T;$$

$$Y = (Y_1,\ldots,Y_n)^T$$

is used. But what is to do if, for fixed x_i, only vague observations $\tilde{Y}_i$ are available? For example, a meteorological station evaluates the clouding $\tilde{Y}$ for a given atmospheric pressure x by expressions like: without clouds, clear, weakly clouded, cloudy, strongly clouded, clouded over. If the vague observations are modelled by fuzzy numbers then the estimator $\hat{\eta}$ of η should also be a fuzzy number. A straightforward approach would be to take the extended LSE

$$\check{\eta} = \check{\lambda}_1 \tilde{Y}_1 \oplus \ldots \oplus \check{\lambda}_n \tilde{Y}_n =: \textcircled{\Sigma} \check{\lambda}_i \tilde{Y}_i \ ; \ \check{\lambda}^T = c^T (F^T F)^{-1} F^T \qquad (4)$$

where $\oplus$ stands for the well-known addition of fuzzy numbers (see, e.g., DUBOIS/PRADE /2/).

In the classical linear regression theory, as the most famous result, the LSE is the best linear unbiased estimator (BLUE) of η. The aim of this paper is to develop a linear estimation theory for linear regression if random fuzzy-set-valued data are available. Especially proposals for definition of expectation and variance of random fuzzy sets are discussed and applied to random fuzzy numbers with stochastically independent random centre and random width (see section 3). In section 4, for a well-defined regression model the BLUE is characterized. The BLUE, in general, does not coincide with the extended LSE (4). Only in special cases the extended LSE can be taken as approximately BLUE.

2. Preliminaries

Let R^d be the d-dimensional Euklidean space. To develop a linear estimation theory with fuzzy observations an addition and scalar multiplication is needed. These operations immediately follow from the extension principle which results in

$$\begin{aligned} m_{A \oplus B}(z) &= \sup_{x+y=z} \inf \{ m_A(x),\ m_B(y) \} \\ m_{\lambda \cdot A}(x) &= m_A(x/\lambda) \ ; \ \lambda \neq 0 \end{aligned} \qquad (5)$$

where m_A, m_B denote the membership functions (abbr.: m.f!s) of the fuzzy sets (abbr.: f.s!s) A and B on R^d. Note that for crisp A and B (5) simplifies to the known MINKOWSKI-addition and to the usual scalar multiplication of crisp sets, i.e.

$$A \oplus B = \{x + y: x \in A,\ y \in B\}$$
$$\lambda \cdot A = \{\lambda x: x \in A\} \tag{6}$$

Using the HAUSDORFF-distance between crisp sets A and B,

$$d(A,B) = \max\{\sup_{x\in A} \inf_{y\in B} \|x-y\| ,\ \sup_{y\in B} \inf_{x\in A} \|x-y\|\} \tag{7}$$

PURI/RALESCU /7/ introduce a HAUSDORFF-distance between f.s!s A, B by

$$\bar{d}(A,B) = \int_0^1 d(A_\alpha,B_\alpha)\, d\alpha \tag{8}$$

where A_α, B_α denote the α-cuts of A and B. Thus, the (compact support) fuzzy subsets $F(R^d)$ on R^d constitute a linear metric space $[F(R^d), \oplus, \cdot, \bar{d}]$.

Especially we are interested in fuzzy numbers, i.e. normalized convex f.s.'s on R^1 which we will use to model vague observations of the type "APPROXIMATELY m". For simplicity we will use LR-type fuzzy numbers with centre m, the left width β and the right width γ, abbreviated by

$$A = [m, \beta, \gamma]_{LR} \tag{9}$$

Then (for details see DUBOIS/PRADE /2/) addition and scalar multiplication simplifies drastically. With $A = [m_A, \beta_A, \gamma_A]_{LR}$, $B = [m_B, \beta_B, \gamma_B]_{LR}$ it holds

$$A \oplus B = [m_A + m_B,\ \beta_A + \beta_B,\ \gamma_A + \gamma_B]_{LR}$$
$$\lambda \cdot A = [\lambda m_A,\ |\lambda|\beta_A,\ |\lambda|\gamma_A]_{LR} \tag{10}$$

3. Expectation and Variance of Random Fuzzy Sets

To introduce random fuzzy sets (abbr. RFS's) we follow the approach by PURI/RALESCU /7/.

Let $(\Omega, \mathcal{B}, P)$ be a probability space and $K(R^d)$ the set of all fuzzy subsets A on R^d with upper semicontinuous, normalized m.f.'s and with compact support. Then a mapping $\tilde{Y}|\Omega \rightarrow K(R^d)$

is called a RFS iff every α-cut $\tilde{Y}_\alpha$ of $\tilde{Y}$ is a compact random set (as defined, e.g., by MATHERON /5/).
In our context we are interested, roughly spoken, in theory and inference on the first and second moment of RFS's.
At first we will consider the **expectation** of a RFS $\tilde{Y}$. With

$$(E\tilde{Y})_\alpha = E\tilde{Y}_\alpha \tag{11}$$

(see PURI/RALESCU /7/) this definition is reduced to the expectation of a (crisp) random set. It is somewhat surprising that in the standard theory of random sets (abbr. RS's) the expectation $E\Xi$ of a RS Ξ is not defined (see e.g. MATHERON /5/).
Using results due to AUMANN /1/ on integrals of set-valued functions, PURI/RALESCU /7/ define

$$E\Xi = \{E\xi : \xi \in \Xi \text{ a.s.} \wedge \xi \in L^1(\Omega, \mathcal{B}, P)\}. \tag{12}$$

Assume that Ω is nonatomic. Then $E\Xi$ is a convex set (see AUMANN /1/) and it holds $E\Xi = E\,co\Xi$, where $co\Xi$ stands for the convex hull of Ξ. Since

$$\begin{aligned} E(\Xi_1 \oplus \Xi_2) &= \{E\xi : \xi \in \Xi_1 \oplus \Xi_2\} = \{E\xi : \xi = \xi_1 + \xi_2,\ \xi_1 \in \Xi_1, \xi_2 \in \Xi_2\} \\ &= \{E\xi_1 + E\xi_2 : \xi_1 \in \Xi_1,\ \xi_2 \in \Xi_2\} = E\Xi_1 \oplus E\Xi_2 \end{aligned} \tag{13}$$

and

$$E\lambda \cdot \Xi = \lambda E\Xi \tag{14}$$

the expectation is a linear operator.
For the case of a deterministic and nonconvex set A we have $EA = coA \neq A$, i.e. the "expectation" EA of a deterministic A does not coincide with A. Thus, (12) is an unsuitable tool to model the "mean value" of nonconvex random figures.

ASS.1: All RS's Ξ and all RFS's $\tilde{Y}$ are assumed to be convex.

For convex RS's in R^1 (random intervals) it holds, with some conditions of measurability (see KRUSE/MEYER /4/)

$$E\Xi = [E \inf \Xi, E \sup \Xi] . \tag{15}$$

Now we return to RFS's. With the appropriate operations $\oplus$ and $\cdot$ from (5) for convex $\tilde{Y}_1$ and $\tilde{Y}_2$ we have, too,

$$\begin{aligned} E(\tilde{Y}_1 \oplus \tilde{Y}_2) &= E\tilde{Y}_1 \oplus E\tilde{Y}_2 \\ E\,\lambda\cdot\tilde{Y} &= \lambda\, E\tilde{Y} \end{aligned} \tag{16}$$

Moreover, for a onedimensional RFS of LR-type, say

$$\tilde{Y} = [m, \beta, \gamma]_{LR} \tag{17}$$

which is convex automatically, we have

$$E\tilde{Y} = [Em, E\beta, E\gamma]_{LR} \tag{18}$$

This easily can be proven using (15) for every α-cut of $\tilde{Y}$. Now let us discuss second moments of RFS's, especially the <u>variance</u>. By

$$(\mathrm{Var}\, \tilde{Y})_\alpha = \mathrm{Var}\, \tilde{Y}_\alpha \tag{19}$$

the variance of a RFS is reduced, also, to the variance of (crisp) RS's. There is a proposal by KRUSE /3/ to define $\mathrm{Var}\,\Xi$ analogously to (12), i.e.

$$\mathrm{Var}\,\Xi = \{\mathrm{Var}\,\xi : \xi \in \Xi \text{ a.s.} \wedge \xi \in L^2(\Omega, \mathcal{B}, P)\} , \tag{20}$$

but this definition has some disadvantages, at least in our context. For instance, according to (20), a deterministic set A has a set-valued variance. E.g. for an interval $A = [-a,a]$ we have: $\mathrm{Var}\,[-a,a] = [0, a^2]$. Thus, (20) measures not only the variability generated by randomness but also, in some sense, the size of the set. In the following we prefer a proposal by STOYAN (private communication) where the variance of a RS Ξ is introduced as the (real-valued) expectation of the squared distance of Ξ from its expectation $E\Xi$, i.e.

$$\mathrm{Var}\,\Xi = E\, d^2(\Xi, E\Xi) . \tag{21}$$

According to (21), clearly, a deterministic convex set A leads to $\operatorname{Var}A = 0$. Let us point out that, in our opinion, a well-founded definition of expectation and variance of RS's from the point of pure probability theory is missing up to now. Thus, (12) and (21) only figure as more or less pragmatic proposals.

For a special case which will be used in the following section we compute the variance of RS's.

<u>Lemma 1</u>: **Let $\Xi = [X \pm \Delta]$ be a random interval with random centre X** and random positive width Δ, both from $L^2(\mathcal{L}, \mathcal{B}, P)$ with variances σ_x^2, σ_Δ^2 and the first absolute central moments $\mu_{|x|}, \mu_{|\Delta|}$. X and Δ are assumed to be stochastically independent. Then

$$\operatorname{Var} \Xi = \sigma_x^2 + \sigma^2 + 2\mu_{|x|}\mu_{|\Delta|} \tag{22}$$

<u>Proof</u>: From (15) directly we have $E\Xi = [EX \pm E\Delta]$. A straight**forward** consideration yields $d(\Xi, E\Xi) = |X-EX| + |\Delta - E\Delta|$. Since X and Δ are independent (22) follows directly from (21).

<u>Remark 1</u>: If Δ is not assumed to be positive we have to consider $\Xi = [X \pm |\Delta|]$.

Unfortunately, for two independent random intervals $\Xi_1 = [X_1 \pm \Delta_1]$ and $\Xi_2 = [X_2 \pm \Delta_2]$ is $\operatorname{Var}(\Xi_1 \oplus \Xi_2) \neq \operatorname{Var}\Xi_1 + \operatorname{Var}\Xi_2$, in some more detail:

$$\begin{aligned}\operatorname{Var}(\Xi_1 \oplus \Xi_2) &= \operatorname{Var}[X_1 + X_2 \pm (\Delta_1 + \Delta_2)] \\ &= \sigma_{x_1}^2 + \sigma_{\Delta_1}^2 + \sigma_{x_2}^2 + \sigma_{\Delta_2}^2 + 2\mu_{|x_1+x_2|}\mu_{|\Delta_1+\Delta_2|}\end{aligned} \tag{23}$$

For scalar multiplication, however, it holds

$$\operatorname{Var} \lambda \cdot \Xi = \lambda^2 \operatorname{Var} \Xi \quad . \tag{24}$$

Returning to RFS's we define with $\bar{d}$ from (8)

$$\operatorname{Var} \tilde{Y} = E\, \bar{d}^2(\tilde{Y}, E\tilde{Y}) \ . \tag{25}$$

<u>Lemma 2</u>: Let $\tilde{Y}$ be a RFS of LR-type with random centre Y, with random width $\beta = \gamma = \Delta > 0$ and with L=R, symbolically

$$\tilde{Y} = [Y, \Delta]_L \tag{26}$$

Furthermore, Y and Δ are assumed to be independent. Then

$$\mathrm{Var}\ \tilde{Y} = \sigma_Y^2 + l^2\, \sigma_\Delta^2 + 2l\, \mu_{|Y|}\mu_{|\Delta|} \quad ; \; l = \int_0^1 |L^{-1}(\alpha)|\, d\alpha \tag{27}$$

<u>Proof</u>: Note that $\tilde{Y}_\alpha = [Y \pm \Delta L^{-1}(\alpha)]$ and (see the proof of Lemma 1): $E\tilde{Y}_\alpha = [EY \pm E\Delta L^{-1}(\alpha)]$;

$d(\tilde{Y}_\alpha, E\tilde{Y}_\alpha) = |Y-EY| + |L^{-1}(\alpha)|\,|\Delta - E\Delta|$;

$\bar{d}(\tilde{Y}, E\tilde{Y}) = |Y-EY| + l\,|\Delta - E\Delta|$

Now, (27) follows from (25) directly ∎

Analogously to (24) we have

$$\mathrm{Var}\ \lambda \cdot \tilde{Y} = \lambda^2\ \mathrm{Var}\ \tilde{Y} \tag{28}$$

4. Linear Regression

As sketched in the **introduction** we consider a regression problem where for fixed (independent) variable x the observation result is a RFS $\tilde{Y}(x)$. To model this situation we assume

$$\tilde{Y}(x) = [Y(x), \Delta]_L \quad ; \quad Y(x) = f(x)^T\vartheta + \varepsilon \tag{29}$$

with known (setup-) function f(x), unknown parameter $\vartheta \in R^r$ and with independent random variables ε and $\Delta > 0$ satisfying

$$E\varepsilon = 0 \ ; \ \mathrm{Var}\,\varepsilon = \sigma_\varepsilon^2 \ ; \ E\Delta = \Delta_o \ ; \ \mathrm{Var}\,\Delta = \sigma_\Delta^2 \ . \tag{30}$$

Applying (18) and Lemma 2 we have

$$E\tilde{Y}(x) = [f(x)^T\vartheta, \Delta_o]_L \tag{31}$$

$$\mathrm{Var}\ \tilde{Y}(x) = \sigma_\varepsilon^2 + l^2\, \sigma_\Delta^2 + 2l\mu_{|\varepsilon|}\mu_{|\Delta|} ; \; l = \int_0^1 |L^{-1}(\alpha)|\, d\alpha \tag{32}$$

For estimation of a linear functional $\eta = c^T\vartheta$ (see (2)) we observe $\tilde{Y}(x)$ stochastically independently at n points $x_1, \ldots, x_n$, i.e.

$$\tilde{Y}(x_i) =: \tilde{Y}_i = [Y_i, \Delta_i]_L \; ; \; Y_i = f(x_i)^T\vartheta + \varepsilon_i \; ; \; i=1,\dots,n \quad (33)$$

We will consider only linear estimators for η, i.e.

$$\hat{\eta} = \lambda_1\tilde{Y}_1 \oplus \dots \oplus \lambda_n\tilde{Y}_n = \textstyle\bigoplus\!\!\sum \lambda_i\tilde{Y}_i \quad ; \quad \lambda_i \in R^1 \quad (34)$$

With (10), (16) and (18) it holds

$$\hat{\eta} = \Big[\sum_{i=1}^{n} \lambda_i Y_i \, , \, \sum_{i=1}^{n} |\lambda_i| \, \Delta_i\Big]_L \quad (35)$$

$$E\hat{\eta} = \Big[\sum_{i=1}^{n} \lambda_i f(x_i)^T\vartheta \, , \, \Delta_o \sum_{i=1}^{n} |\lambda_i| \Big]_L \quad (36)$$

An estimator $\hat{\eta}$ is called unbiased iff the centre of the fuzzy expectation (36) coincides with the unknown $\eta = c^T\vartheta$, i.e. iff

$$\sum_{i=1}^{n} \lambda_i f(x_i) = F^T\lambda = c \quad . \quad (37)$$

Our aim is to find the BLUE for η. With Lemma2 and the assumed independence of the observations we obtain

$$\begin{aligned} \mathrm{Var}\,\hat{\eta} &= \sigma^2_{\Sigma\lambda_i\varepsilon_i} + 1^2\, \sigma^2_{\Sigma|\lambda_i|\Delta_i} + 21\mu_{|\Sigma\lambda_i\varepsilon_i|}\mu_{|\Sigma|\lambda_i|\Delta_i|} \\ &= \lambda^T\lambda\big[\sigma^2_\varepsilon + 1^2\,\sigma^2_\Delta\big] + 21\mu_{|\Sigma\lambda_i\varepsilon_i|}\mu_{|\Sigma|\lambda_i|\Delta_i|} =: \varphi(\lambda) \end{aligned} \quad (38)$$

Thus, we have proved

Theorem 1: The BLUE $\hat{\eta}^*$ for $\eta = c^T\vartheta$ is given by $\hat{\eta}^* = \textstyle\bigoplus\!\!\sum \lambda_i^* \tilde{Y}_i$ where $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$ is the solution of $\varphi(\lambda) = \inf$ w.r.t. the side condition $F^T\lambda = c$.

Due to the dependence from absolute moments, which do not follow a simple law by addition of random variables, the functional $\varphi(\lambda)$, which is to be minimized, is highly complicated. Thus, our linear estimation theory does not share the elegancy known from classical regression case.

In the following we will show that the extended LSE (4) appears as an approximate solution in a special case. We suppress, however, all quantitative terms and present only qualitative arguments.

<u>ASS.2</u>: ε and Δ are Gaussian variables.

Since Δ can be negative now we really have to work with $|\Delta|$ (see Remark 1). We assume, however, that Δ only with a small probability is negative (e.g. if $\Delta_0 \geq 3\sigma_\Delta$). Thus, qualitatively we have $\Delta \approx |\Delta|$. For $\tilde{Y}(x)$ from (29) then it holds approximately

$$\mathrm{Var}\ \tilde{Y}(x) \approx \sigma_\varepsilon^2 + 1^2\ \sigma_\Delta^2 + 41\sigma_\varepsilon\,\sigma_\Delta/\pi \ . \tag{39}$$

This follows from (32) and the fact that with ASS.2

$$\mu_{|\varepsilon|} = \sqrt{2/\pi}\ \sigma_\varepsilon \quad ; \quad \mu_{|\Delta|} = \sqrt{2/\pi}\ \sigma_\Delta \ . \tag{40}$$

Since $\sum_{i=1}^{n}\lambda_i\varepsilon_i$ and $\sum_{i=1}^{n}|\lambda_i|\Delta_i$ are Gaussian variables we obtain, analogously to (40)

$$\mu_{|\Sigma\lambda_i\varepsilon_i|} = \sqrt{2/\pi}\,\sigma_{\Sigma\lambda_i\varepsilon_i} = \sqrt{2\,\lambda^T\lambda/\pi}\ \sigma_\varepsilon$$

$$\mu_{|\Sigma|\lambda_i|\Delta_i|} = \sqrt{2/\pi}\,\sigma_{\Sigma|\lambda_i|\Delta_i} = \sqrt{2\,\lambda^T\lambda/\pi}\ \sigma_\Delta \ .$$

Thus, for $\hat{\eta}$ from (34) the variance (38) reduces (approximately) to

$$\mathrm{Var}\ \hat{\eta} \approx \lambda^T\lambda\ \mathrm{Var}\ \tilde{Y}(x) \ . \tag{41}$$

Minimization of (41) w.r.t. (37), i.e.

$$\lambda^T\lambda = \min \quad ; \quad F^T\lambda = c \tag{42}$$

leads to the well-known LSE-coefficient vector $\check{\lambda}$ in (4). Thus, the extended LSE (4) appears as approximately BLUE for $\eta = c^T\vartheta$ if ASS.2 and $\Delta \approx |\Delta|$ are satisfied. In general, however, the extended LSE is not BLUE (in the sense of Theorem 1) in an estimation theory of second order for the considered regression problem.

A simple example for the extended LSE and further remarks on minimum width unbiased linear estimators for η can be found in NÄTHER/ALBRECHT /6/.

5. References

/1/ AUMANN, R.J.: Integrals of set-valued functions, J.Math.Anal.Appl. 12 (1965), 1-12

/2/ DUBOIS, D. and H. PRADE: Fuzzy Sets and Systems: Theory and Application, Academic Press, New York 1980

/3/ KRUSE, R.: On the Variance of Random Sets, J.Math.Anal.Appl. 122 (1987), 469-473

/4/ KRUSE, R. and K.D. MEYER: Statistics with vague data, Reidel Publ. Comp., Dordrecht 1987

/5/ MATHERON, G.: Random Sets and Integral Geometry, Wiley, New York 1975

/6/ NÄTHER, W. and M. ALBRECHT: Linear Regression with Random Fuzzy Observations, submitted to Statistics

/7/ PURI, M.L. and D. RALESCU: Fuzzy Random Variables, J.Math.Anal.Appl. 114 (1986), 409-422.