Regression with Evidential Coefficients

Alexander Lepskiy

National Research University – Higher School of Economics,
Moscow, Russia

INFUS 2025, July 29–31, 2025, Istanbul, Türkiye

Research Motivation

There are many techniques for estimating the regression function:

- methods of mathematical statistics (e. g., the method of least squares);
- machine learning methods:
  - K-nearest neighbor smoother,
  - SVM regression,
  - regularization-based methods (ridge regression, Lasso method),
  - splines, etc.

Classical regression methods assume that the data sources are **reliable**, and the data itself is **accurate**.

But not all data is reliable and accurate!

Therefore, the problem of regression analysis with **imprecise** and **uncertain** (unreliable) data is relevant.

Possible modelling methods:

- inaccuracies — fuzzy sets (fuzzy regression: [Tanaka 1987, Diamond 1988, etc.];
- uncertainties — evidence theory (EVidential REGression [Petit-Renaud & Denœux 2004], Evidential Neural Network regression [Denœux 2023]).

The latest methods are implemented based on the K-nearest neighbor method and belongs to the group of local (nonparametric) regression analysis methods.

A new approach will be proposed that develops the possibilistic model [Tanaka 1987] for finding fuzzy linear regression coefficients. At the same time, this new model of **Evidential Regression** (ER) will take into account information about the degree of belief in the found imprecise (interval or fuzzy) regression coefficients within the framework of evidence theory [Dempster 1967, Shafer 1974].

Outline of Presentation

- Background from the Theory of Evidence;
- Statement of the ER Problem;
  - ER with Interval Coefficients;
  - Conjunctive Aggregation of Jointly Consonant Bodies of Evidence;
  - ER with Fuzzy Coefficients;
- Numerical Example;
- Summary and Conclusion.

Background from the Theory of Evidence

[Dempster 1967, Shafer 1974]

Let:

- $X \subseteq \mathbb{R}^n$, 2^X be the set of all subsets of X;
- $\mathcal{B}$ is a finite set of subsets of X (focal elements);
- $m : 2^X \to [0, 1]$ is a mass function, $m(B) > 0 \Leftrightarrow B \in \mathcal{B}$, $\sum_{B\in\mathcal{B}} m(B) = 1$;
- $F = (\mathcal{B}, m)$ is called a **body of evidence** (**BE**);
    - the BE $F_B = (\{B\}, 1)$ with one focal element is called categorical;
    - the BE F_X is a vacuous;
- the BE $F = (\mathcal{B}, m)$ can be represented as $F = \sum_{B\in\mathcal{B}} m(B)F_B$;
- the BE $F_B^\beta = (1 - \beta)F_B + \beta F_X$, $\beta \in [0, 1]$ is called simple BE;
- the BE $F = (\mathcal{B}, m)$ is called consonant if $B' \subseteq B''$ or $B'' \subseteq B'$ is true for $\forall B', B'' \in \mathcal{B}$.

The **belief** and **plausibility** functions

$$Bel(B) = \sum_{C \subseteq B} m(C), \;\; Pl(B) = \sum_{B \cap C \neq \emptyset} m(C)$$

are assigned to the BE $F = (\mathcal{B}, m)$. The function

$$Pl(x) = \sum_{C : x \in C} m(C)$$

is called the **contour function**. For BE with normal fuzzy focal elements this function is equal to

$$Pl(x) = \sum_{\tilde{C}} m(\tilde{C}) \mu_{\tilde{C}}(x).$$

The contour function coincides with the possibility distribution function (= membership function of a fuzzy set) for consonant BE.

The **degree of uncertainty** of the BE $F = (\mathcal{B}, m)$ is characterized using the functional

$$H(F) = \sum_{B \in \mathcal{B}} m(B)\lambda(B),$$

where λ is the Lebesgue measure. If $B = \tilde{B}$ is a fuzzy set, then

$$\lambda(\tilde{B}) = \int_X \mu_{\tilde{B}}(t)\lambda(dt),$$

where $\mu_{\tilde{B}}$ is the membership function.

We will use the **conjunctive rule of combination** $F = (\mathcal{B}, m) = \bigotimes_{k=1}^{l} F_k$ of BEs $F_k = (\mathcal{B}_k, m_k)$, $k = 1, \ldots, l$ according to the rule:

$$m(B) = \sum_{\substack{B_1 \cap \ldots \cap B_l = B, \\ B_k \in \mathcal{B}_k, k=1,\ldots,l}} m_1(B_1) \cdot \ldots \cdot m_l(B_l). \tag{1}$$

For non-conflicting BEs (i. e. $B_1 \cap \ldots \cap B_l \neq \emptyset \ \forall B_k \in \mathcal{B}_k$, $k = 1, \ldots, l$) this rule coincides with **Dempster's rule** [Dempster 1967].

We will consider Ishizuka's [Ishizuka 1982] approach of generalizing Dempster's rule to the case of BEs with fuzzy focal elements:

$$m(\tilde{B}) = \frac{1}{k} \sum_{\substack{\tilde{B}_1 \cap \ldots \cap \tilde{B}_l = \tilde{B}, \\ \tilde{B}_k \in \mathcal{B}_k, k=1,\ldots,l}} h_{\tilde{B}_1 \cap \ldots \cap \tilde{B}_l} m_1(\tilde{B}_1) \cdot \ldots \cdot m_l(\tilde{B}_l), \tag{2}$$

where $k = \sum_{\tilde{B}_k \in \mathcal{B}_k, k=1,\ldots,l} h_{\tilde{B}_1 \cap \ldots \cap \tilde{B}_l} m_1(\tilde{B}_1) \cdot \ldots \cdot m_l(\tilde{B}_l)$, $h_{\tilde{B}} = \sup_x \mu_{\tilde{B}}(x)$.

Statement of the ER Problem

We consider the problem of approximating point data $\{(x_i, y_i)\}_{i=1}^N$ by a function $f(x; A_0, \ldots, A_n)$, where A_j, $j = 0, \ldots, n$ are BEs. For simplicity, we will further consider only the case of paired linear regression:

$$f(x; A_0, A_1) = A_0 + A_1 x.$$

We will assume that ER-coefficients are simple BEs of the form

$$A_j^\alpha = (1 - \alpha) F_{[a_j, b_j]} + \alpha F_{X_j},$$

where $X_j \subseteq \mathbb{R}$, $j = 0, 1$.

Each ER-coefficient is determined by one focal element (an interval or a fuzzy number) and the degree of confidence that the true value of the coefficient belongs to this element.

ER with Interval Coefficients

Note that specifying a pair of simple BEs-coefficients is equivalent to specifying one simple BE on the set of parameters

$$F_D^{\alpha} = (1-\alpha)F_D + \alpha F_{\Pi}, \tag{3}$$

where D (Π) are some rectangle in $\mathbb{R}^2$, $D \subseteq \Pi$ limited by lines with the desired (possible) parameters:

$$D = \{\mathbf{w} = (w_0, w_1) : L_{\mathbf{w}}(x) = w_0 + w_1 x \approx \{(x_i, y_i)\}_{i=1}^{N}\}.$$

The model will consist of the following steps.

1. To determine Π we find the regression line $L_{\mathbf{c}}(x) = c_0 + c_1 x$ (assuming that the errors are distributed according to the normal law $N(0, \sigma^2)$) using the formulas:

$$c_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \;\; c_0 = \bar{y} - c_1 \bar{x},$$

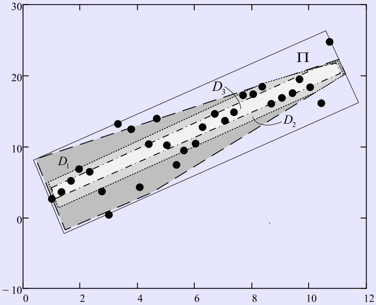
$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \; \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i.$$

   Next, we find the rectangle Π containing all sample points and located "along" the regression line $L_{\mathbf{c}}$.

2. We find two lines of the form $L^{\pm}(x) = (c_0 \pm \Delta_0) + (c_1 \pm \Delta_1)x$ which, together with the rectangle Π, limit the domain $D_s \subseteq \Pi$ of minimal area and satisfying the condition

$$\frac{1}{N} \left| \{ i : (x_i, y_i) \in D_s \} \right| \geq 1 - \alpha_s, \;\; \alpha_s \in (0, 1).$$

3. The problem of step 2 is solved for l values $0 < \alpha_1 < \ldots < \alpha_l \leq 1$. As a result, we obtain l simple BEs $F_{D_k}^{\alpha_k}$, $k = 1, \ldots, l$.



4. Simple BEs $F_{D_k}^{\alpha_k}$, $k = 1, \ldots, l$ are aggregated using the conjunctive rule $F = \otimes_{k=1}^{l} F_{D_k}^{\alpha_k}$. As a result, we will obtain the final BE, which will determine the ER coefficients.

Conjunctive Aggregation of Jointly Consonant BEs

The simple BEs $F_{D_k}^{\alpha_k}$, $k = 1, \dots, l$ may turn out to be **jointly consonant**, i. e. there exists a permutation of indices such that: $D_{i_1} \subseteq \dots \subseteq D_{i_l} \subseteq \Pi$.

Proposition

The BE $F = (\mathcal{A}, m) = \bigotimes\limits_{k=1}^{l} F_{D_k}^{\alpha_k}$ obtained as a result of conjunctive aggregation of jointly consonant simple BEs $F_{D_k}^{\alpha_k}$, $k = 1, \dots, l$ will be consonant, $\mathcal{A} = \{D_1, \dots, D_l, \Pi\}$. If $\emptyset = D_{i_0} \subseteq D_{i_1} \subseteq \dots \subseteq D_{i_l} \subseteq D_{i_{l+1}} = \Pi$, then

$$m\left(D_{i_k}\right) = \left(1 - \alpha_{i_k}\right) \alpha_{i_0} \alpha_{i_1} \dots \alpha_{i_{k-1}},\ k = 1, \dots, l,\ m\left(\Pi\right) = \alpha_{i_1} \dots \alpha_{i_l},$$

$$Pl(x) = \alpha_{i_0} \dots \alpha_{i_{k-1}} \text{ if } x \in D_{i_k} \backslash D_{i_{k-1}},\ k = 1, \dots, l+1,\ \alpha_{i_0} = 1.$$

ER with Fuzzy Coefficients

Let us now consider the case when the ER-coefficients are triangular fuzzy numbers of the form $\tilde{a}_j = (c_j - \Delta_j, c_j, c_j + \Delta_j)$, $\Delta_j \geq 0$, $j = 0, 1$. Then the information that the ER-coefficients are such fuzzy numbers with belief level $1 - \alpha$ can be represented by a simple BE

$$A_j^\alpha = (1-\alpha)F_{(c_j - \Delta_j, c_j, c_j + \Delta_j)} + \alpha F_{X_j},\ j = 0, 1.$$

Let us set the problem of finding triangular fuzzy numbers (parameters c_j, $\Delta_j \geq 0$) for which:

1) the uncertainty functional $\Phi(H(A_0), H(A_1)) \to \min$ (Φ is the convolution of two terms; e.g. $\Phi = H(A_0) + \lambda H(A_1)$;
2) at least $(1-\alpha)N$ sample points would fall into a given h-cut of the model solution $(\tilde{a}_0)_h + (\tilde{a}_1)_h x$.

The last requirement is equivalent to the condition

$$\left|\{i : |c_0 + c_1 x_i - y_i| \leq (1-h)\left(\Delta_0 + \Delta_1 |x_i|\right)\}\right| \geq (1-\alpha)N, \tag{4}$$

If we use the measure of the uncertainty $H(A) = \sum_{\tilde{a}\in\mathcal{A}} m(\tilde{a})\,|\tilde{a}|$ for the BE $A = \sum_{\tilde{a}\in\mathcal{A}} m(\tilde{a}) F_{\tilde{a}}$, then

$$H(A_j^\alpha) = (1-\alpha)\Delta_j + \alpha\,|X_j| \propto \Delta_j \;\; \text{(for fixed } \alpha, X_j\text{)}.$$

In this case $\Phi(H(A_0^\alpha), H(A_1^\alpha)) \to \min \;\; \Leftrightarrow \;\; \tilde{\Phi}(\Delta_0, \Delta_1) \to \min$. Then the function to be minimized takes the form

$$\sum_{i\in I} \left(c_0 + c_1 x_i - y_i\right)^2 + \theta\tilde{\Phi}(\Delta_0, \Delta_1) \text{ under condition (4)},$$

where $\theta > 0$, $I = \{i : |c_0 + c_1 x_i - y_i| \leq (1-h)\left(\Delta_0 + \Delta_1 |x_i|\right)\}$. The first term (MSE) is added to obtain a stable solution.

If we solve this problem for l values $0 < \alpha_1 < \ldots < \alpha_l \leq 1$, we obtain l pairs of simple BEs $A_0^{\alpha_k}$, $A_1^{\alpha_k}$, $k = 1, \ldots, l$ with fuzzy focal elements, which can then be aggregated using the fuzzy conjunctive rule.

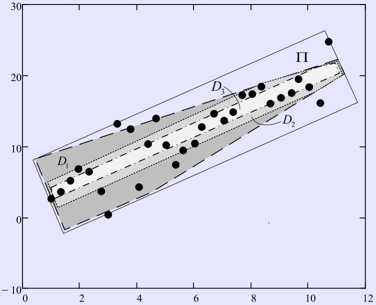
As a result, we obtain ER-coefficients $A_j = \bigotimes_{k=1}^{l} A_j^{\alpha_k}$, $j = 0, 1$.

Numerical Example. ER with Interval Coefficients

Let us present the results of ER on synthetic data $\{(x_i, y_i)\}_{i=1}^{30}$:

$$x_i \sim N(\tfrac{1}{3}i + 1, 0.0009), \quad i = 1, \dots, 30. \qquad y_i \sim \begin{cases} N(1 + 2x_i, 16), & i = 1, \dots, 10, \\ N(1 + 2x_i, 4), & i = 11, \dots, 30. \end{cases}$$

For three values $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, $\alpha_3 = 0.5$, we find the optimal domains $\Pi = D_0 \supseteq D_1 \supseteq D_2 \supseteq D_3$, the boundaries of which correspond to the boundary values of the intervals of the ER-coefficients.



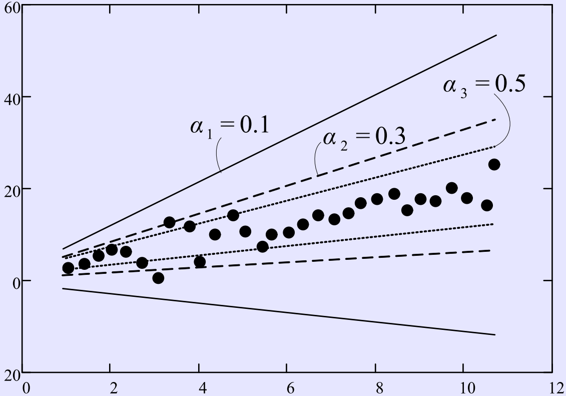
As a result, we obtain jointly consonant BEs, from which a contour function can be constructed:

$$Pl(x,y) = \begin{cases} 1, & (x,y) \in D_3, \\ 0.5, & (x,y) \in D_2 \backslash D_3, \\ 0.15, & (x,y) \in D_1 \backslash D_2, \\ 0.015, & (x,y) \in \Pi \backslash D_1. \end{cases}$$

This function can be viewed as a membership function of a fuzzy set that determines the ER-coefficients.

Numerical Example. ER with Fuzzy Coefficients

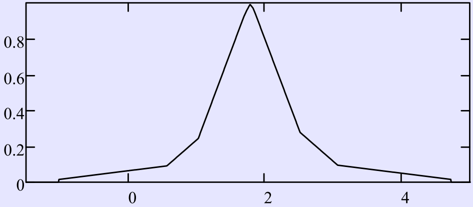
The results of ER with fuzzy coefficients (only the support boundaries are visualized) for values $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, $\alpha_3 = 0.5$ and $h = 0.7$, $\Phi(t, s) = 0.6t + 0.4s$ are shown in Fig.



The coefficients are triangular fuzzy numbers $\tilde{a}_0$ and $\tilde{a}_1$, the parameters of which for three values α are given in Table.

	$\tilde{a}_0$	$\tilde{a}_1$
$\alpha_1 = 0.1$	$(-0.89,\ 1.02,\ 2.94)$	$(-1.02,\ 1.85,\ 4.72)$
$\alpha_2 = 0.3$	$(0.62,\ 1.45,\ 2.28)$	$(0.56,\ 1.81,\ 3.06)$
$\alpha_3 = 0.5$	$(1.38,\ 1.88,\ 2.37)$	$(1.02,\ 1.76,\ 2.51)$

We will perform aggregation $A_j = \bigotimes_{k=1}^{3} A_j^{\alpha_k}$, $j = 0, 1$ of simple BE-coefficients using Dempster's rule for fuzzy focal elements. As a result, we will obtain the final ER-coefficients. The graph of the contour function Pl_1 of the ER-coefficient A_1 is shown in Fig.



Summary and Conclusion

Two models of ER were considered: with interval and fuzzy focal elements.

The advantages of these models are:

- new models allow us to obtain more complete information about the desired coefficients with a lower degree of blurring compared to fuzzy regression models;
- these models will be more robust compared to possibility models of fuzzy regression, since the method does not require that all sample elements (including outliers) belong to a given cutting set.

Generalization of the proposed methodology to the case of multiple regression (including the development of computational procedures) is an important problem for further research.

References

- Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. **38**, 325–339 (1967)
- Denœux, T.: Quantifying prediction uncertainty in regression using random fuzzy sets: the ENNreg model. IEEE Trans. on Fuzzy Syst. **31**(10), 3690–3699 (2023)
- Diamond, P.: Fuzzy least squares. Information Science **46**, 141–157 (1988)
- Ishizuka, M.: An extension of Dempster & Shafer's theory to fuzzy set for constructing expert systems. Seisan-Kenkyu **34**(7), 312–315 (1982)
- Petit-Renaud, S., Denœux, T.: Nonparametric regression analysis of uncertain and imprecise data using belief functions. Intern. J. of Approx. Reas. **35**(1), 1–28 (2004)
- Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press (1976)
- Tanaka, H.: Fuzzy data analysis by possibilistic linear models. Fuzzy Sets and Syst. **24**, 363–375 (1987)

Thanks for you attention

alepskiy@hse.ru
https://www.hse.ru/en/org/persons/10586209