

A CONSISTENT CONFIDENCE INTERVAL FOR FUZZY CAPABILITY INDEX

A. PARCHAMI †, M. MASHINCHI †, V. PARTOVI NIA ‡, §

ABSTRACT. Fuzzy process capability indices are used to determine whether a production process is capable of producing items within specification tolerance, where instead of precise quality we have two membership functions for specification limits. In practice these indices are estimated using sample data and it is of interest to obtain confidence limits for fuzzy capability index given a sample estimate. After introducing $100(1 - \alpha)\%$ fuzzy confidence interval by Parchami et al. in 2005 and 2006, for fuzzy capability index \tilde{C}_p , our observation leads us to propound an open problem about the consistency property for interval estimation of \tilde{C}_p . In this paper we redefine some concepts about fuzzy confidence interval and then we show the consistency property of the fuzzy confidence interval proposed by Parchami et al. in 2006 holds for almost every α in $[0, 1]$.

Keywords: fuzzy process capability index, fuzzy confidence interval, triangular fuzzy number, weak law of large numbers.

1. PRELIMINARIES

When we use the precise specification limits, several statistics such as C_p , C_{pm} , C_{pk} and C_{pk} are used to estimate the capability of a manufacturing process, which in most cases it is assumed we have a large sample from a normal population [6]. If we introduce vagueness into specification limits, we face quite new and interesting problems and the classical capability indices could not be applied. For such cases Yongting [18] introduced a process capability index C_p as a real number and it was used by Sadeghpour-Gildeh [16]. Lee investigated a process capability index, C_{pk} , as a fuzzy set [7]. Parchami et al. introduced fuzzy process capability indices as fuzzy numbers and discussed relations that governing between them when specification limits are fuzzy rather than crisp [11, 12, 14]. The organization of this paper is as follows. In Section 2, we review traditional and fuzzy capability indices and then we review ranking functions in Section 3. In Section 4, we reintroduce a fuzzy confidence interval for fuzzy capability index \tilde{C}_p and then we prove an open problem presented by Parchami at al. in [11, 14]. A conclusion presented in the final section.

Let \mathbb{R} be the set of real numbers. Assume $F(\mathbb{R})$ be the set of all real valued continuous functions from \mathbb{R} to $[0, 1]$, i.e. $F(\mathbb{R}) = \{\tilde{A}|\tilde{A} : \mathbb{R} \rightarrow [0, 1], \tilde{A} \text{ is a continuous function}\}$. Also suppose that

$$F_T(\mathbb{R}) = \{T_{a,b,c} | a, b, c \in \mathbb{R}, a \leq b \leq c\},$$

where

$$T_{a,b,c} = \begin{cases} (x - a)/(x - b) & \text{if } a \leq x < b, \\ (c - x)/(c - b) & \text{if } b \leq x < c, \\ 0 & \text{elsewhere.} \end{cases} \quad (1)$$

†Department of Statistics, Faculty of Mathematics and Computer Sciences, Shahid Bahonar University of Kerman, Kerman, Iran, fax: 0098-341-3221080, E-mail: mashinchi@mail.uk.ac.ir.

‡Ecole Polytechnique Fédérale de Lausanne (EPFL) CH-1015 Lausanne, Switzerland

§*Manuscript received 15 February, 2008.*

Any $\tilde{A} \in F(\mathbb{R})$ is called a fuzzy set on \mathbb{R} and any $T_{a,b,c} \in F_T(\mathbb{R})$ is called a triangular fuzzy number, which we sometimes write as $T(a, b, c)$. We assume $T(a, a, a)$ be $I_{\{a\}}$, the indicator function of a . The following definition could be given by using the extension principle [10].

Definition 1. Let $T(a, b, c) \in F_T(\mathbb{R})$, $k \in \mathbb{R}$ and $k \geq 0$. The operation \otimes on $F_T(\mathbb{R})$ is defined as follows

$$k \otimes T(a, b, c) = T(a, b, c) \otimes k = T(ka, kb, kc). \quad (2)$$

2. PROCESS CAPABILITY INDEX

2.1. Traditional Capability Index. Process capability compares the output of a process with the specification limits by using capability indices. Frequently, this comparison is made by forming the ratio of the width between the process specification limits with the width of the natural tolerance limits, as measured by 6 process standard deviation units. This method leads to making a statement about how well the process meets specifications [6]. For convenience, we will denote the upper and lower limits respectively by U and L , rather than using the more customary USL and LSL notations. We deal with univariate measurements, where its corresponding random variable X will have mean μ and standard deviation σ . We assume that the measured characteristic should have or at least approximately normal distribution [6].

The commonly recognized process capability index is

$$C_p = \frac{U - L}{6\sigma}, \quad (3)$$

which is ascribed to Juran [4] and it is used when $\mu = \frac{U+L}{2}$. Substituting the standard deviation in (3) will provide a point estimate for this index. We would never expect this point estimate to be exactly equal to the real value of the population parameter. So we often compute a $100(1 - \alpha)\%$ confidence interval for our parameter. In practice a confidence bound can be used to guard against false optimism. Kane [5] suggested the $100(1 - \alpha)\%$ confidence interval limits of C_p as follows

$$\left[\hat{C}_p \sqrt{\frac{\chi_{n-1, \alpha/2}^2}{n-1}}, \hat{C}_p \sqrt{\frac{\chi_{n-1, 1-\alpha/2}^2}{n-1}} \right] \quad (4)$$

where $\hat{C}_p = \frac{U-L}{6s}$, $s = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2}$ and $\chi_{n-1, \alpha}^2$ is the α -quantile of chi-square with $n - 1$ degrees of freedom.

2.2. Fuzzy Capability Index. When we have fuzzy specification limits, it will be more realistic to have a C_p which is also fuzzy. This is due to the fact that a fuzzy capability index could be more informative than a precise number. For this situation Parchami et al. [14] introduced capability indices as fuzzy numbers. Suppose that $U(a_u, b_u, c_u) = T(a_u, b_u, c_u) \in F_T(\mathbb{R})$ and $L(a_l, b_l, c_l) = T(a_l, b_l, c_l) \in F_T(\mathbb{R})$ be the engineering fuzzy specification limits, where $a_u \geq c_l$. Then the new fuzzy process capability index is defined as follows

$$\tilde{C}_p = T\left(\frac{a_u - c_l}{6\sigma}, \frac{b_u - b_l}{6\sigma}, \frac{c_u - a_l}{6\sigma}\right). \quad (5)$$

Note that \tilde{C}_p is useful when $\mu = \frac{b_u + b_l}{2}$. A generalized version of the above fuzzy capability index is discussed in [9], where the specification limits are LR fuzzy intervals.

3. RANKING FUNCTION

In the next section we are going to define a fuzzy confidence interval for \tilde{C}_p , where comparing fuzzy numbers is emergent and so an ordering approach is needed. A simple but efficient approach for the ordering of the elements of $F(\mathbb{R})$ is to define a ranking function $R : F(\mathbb{R}) \rightarrow \mathbb{R}$

which maps each fuzzy number into the real line, where a natural order exists [8]. Define the order \leq_R on $F(\mathbb{R})$ by

$$\begin{aligned} \tilde{A} \geq_R \tilde{B} & \text{ if and only if } R(\tilde{A}) \geq R(\tilde{B}), \\ \tilde{A} \leq_R \tilde{B} & \text{ if and only if } R(\tilde{A}) \leq R(\tilde{B}), \\ \tilde{A} =_R \tilde{B} & \text{ if and only if } R(\tilde{A}) = R(\tilde{B}), \end{aligned}$$

where \tilde{A} and \tilde{B} are in $F(\mathbb{R})$. Several ranking functions R have been proposed by researchers to suit their requirements of the problems under consideration. The ranking function proposed by Roubens [3, 15] is defined by

$$R_r(\tilde{A}) = \frac{1}{2} \int_0^1 \left(\inf \tilde{A}_\alpha + \sup \tilde{A}_\alpha \right) d\alpha. \tag{6}$$

Lemma 1. *If $T(a, b, c) \in F_T(\mathbb{R})$, then Roubenss ranking function reduces to*

$$R_r(T(a, b, c)) = \frac{2b + a + c}{4}.$$

Proof. See Lemma 4.1 of [11].

Lemma 2. *Let $m, n \in \mathbb{R}$, $T(a, b, c) \in F_T(\mathbb{R})$ and $2b + a + c \geq 0$. Then according to Roubenss ranking function we have*

$$m \otimes T(a, b, c) \leq n \otimes T(a, b, c) \text{ if and only if } m \leq n.$$

Proof. See Lemma 4.2 of [11].

Lemma 3. *Let $k \in \mathbb{R}$ and $T(a, b, c) \in F_T(\mathbb{R})$. Then according to Roubenss ranking function*

$$R_r(k \otimes T(a, b, c)) = k R_r(T(a, b, c)).$$

Proof. Proof is obvious, by using Definition 1 and Lemma1

4. FUZZY CONFIDENCE INTERVAL FOR \tilde{C}_p

Substituting the standard deviation in (5) provides a fuzzy point estimate for \tilde{C}_p . This point estimate of \tilde{C}_p is denoted by $\hat{\tilde{C}}_p$. Since $\hat{\tilde{C}}_p$, like other statistics, is subject to sampling variation, it is critical to compute a confidence interval to provide a range which includes the true \tilde{C}_p with high probability.

Definition 2. [11] *Let $\tilde{A}, \tilde{B} \in F_T(\mathbb{R})$ and $\tilde{A} \leq_R \tilde{B}$. The fuzzy interval $[\tilde{A}, \tilde{B}]$ is the set*

$$[\tilde{A}, \tilde{B}] = \left\{ \tilde{C} \in F_T(\mathbb{R}); \tilde{A} \leq_R \tilde{C} \leq_R \tilde{B} \right\}.$$

Suppose that the set of all random samples of size n which are possible is $X^{(n)}$.

Definition 3. *Any function $\tilde{A} : X^{(n)} \rightarrow F_T(\mathbb{R})$ is called a fuzzy statistic. Note that $\tilde{A}(X_1, \dots, X_n)$ depends only on the random sample X_1, \dots, X_n and not any unknown parameters. When the observation $\mathbf{x} = (x_1, \dots, x_n)$ is given, then the value of the statistic $\tilde{A}(\mathbf{x})$ is just one triangular fuzzy number.*

Let X be a measurable random variable on the probability space $(\Omega, \mathcal{F}, \Pr)$ and $\tilde{T} = T(a, b, c) \in F_T(\mathbb{R})$ be such that $2b + a + c \geq 0$. According to (2), for any $\omega \in \Omega$ we define $(X \otimes \tilde{T})(\omega) = X(\omega) \otimes \tilde{T}$.

Lemma 4. Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \Pr)$, $k_1, k_2 \in \mathbb{R}$ and $\tilde{T} = T(a, b, c) \in F_T(\mathbb{R})$, where $2b + a + c \geq 0$. Then

$$\Pr(k_1 \otimes \tilde{T} \leq X \otimes \tilde{T} \leq k_2 \otimes \tilde{T}) = 1 - \alpha \quad \text{if and only if} \quad \Pr(k_1 \leq X \leq k_2) = 1 - \alpha.$$

Proof. See Proposition 5.1 of [11].

According to the Lemma 4, we can give the following definition.

Definition 4. Let $\tilde{A}, \tilde{B} \in F_T(\mathbb{R})$ be the observed fuzzy statistic, where $\tilde{A} \leq \tilde{B}$. Then $[\tilde{A}, \tilde{B}]$ is called a $100(1 - \alpha)\%$ fuzzy confidence interval for $X \otimes \tilde{T}$, where $\Pr(\tilde{A} \leq X \otimes \tilde{T} \leq \tilde{B}) = 1 - \alpha$.

Theorem 1. Suppose X_1, X_2, \dots, X_n that are independent, identically distributed random variables with $N(\mu, \sigma^2)$ and $U(a_u, b_u, c_u) \in F_T(\mathbb{R})$, $L(a_l, b_l, c_l) \in F_T(\mathbb{R})$ are the engineering fuzzy specification limits, where $a_u \geq c_l$. Then the following interval is a $100(1 - \alpha)\%$ fuzzy confidence interval for \hat{C}_p

$$\left[\hat{C}_p \otimes \sqrt{\frac{\chi_{n-1, \alpha/2}^2}{n-1}}, \hat{C}_p \otimes \sqrt{\frac{\chi_{n-1, 1-\alpha/2}^2}{n-1}} \right] \tag{7}$$

where $\hat{C}_p = T\left(\frac{a_u - c_l}{6s}, \frac{b_u - b_l}{6s}, \frac{c_u - a_l}{6s}\right)$ is the point estimation of \tilde{C}_p .

Proof. See Theorem 5.1 of [11].

Definition 5. Let $[\tilde{A}, \tilde{B}]$ and $[\tilde{A}_n, \tilde{B}_n]$; $n \in \mathbb{N}$ be fuzzy intervals. Define

$$i) \quad [\tilde{A}_1, \tilde{B}_1] = [\tilde{A}_2, \tilde{B}_2] \quad \text{if} \quad [R(\tilde{A}_1), R(\tilde{B}_1)] = [R(\tilde{A}_2), R(\tilde{B}_2)], \tag{8}$$

$$ii) \quad \lim_{n \rightarrow \infty} [\tilde{A}_n, \tilde{B}_n] = [\tilde{A}, \tilde{B}] \quad \text{if} \quad \lim_{n \rightarrow \infty} [R(\tilde{A}_n), R(\tilde{B}_n)] = [R(\tilde{A}), R(\tilde{B})], \tag{9}$$

where $\lim_{n \rightarrow \infty} [R(\tilde{A}_n), R(\tilde{B}_n)] = \lim_{n \rightarrow \infty} R(\tilde{A}_n), \lim_{n \rightarrow \infty} R(\tilde{B}_n)$.

Our observation leads us to the following theorem which was presented as an open problem in [11, 13]. Now we are ready to give the main result of this paper.

Theorem 2. Assuming $m_{[0,1]}$ is the Lebesgue measure defined on $[0, 1]$. Under the same assumption as in Theorem 1, we have

$$m_{[0,1]} \left(\left\{ \alpha; \lim_{n \rightarrow \infty} \left[\hat{C}_p \otimes \sqrt{\frac{\chi_{n-1, \alpha/2}^2}{n-1}}, \hat{C}_p \otimes \sqrt{\frac{\chi_{n-1, 1-\alpha/2}^2}{n-1}} \right] = \left[\hat{C}_p, \hat{C}_p \right] \right\} \right) = 1. \tag{10}$$

Proof. Since any χ_n^2 random variable may be written as the sum of n i.i.d. χ_1^2 ; so by the weak law of large numbers (Theorem 6.4.3 of [1]), we have

$$\lim_{n \rightarrow \infty} \Pr \left(\left| \frac{\chi_n^2}{n} - 1 \right| < \varepsilon \right) = 1 \quad \text{for every} \quad \varepsilon > 0. \tag{11}$$

Since $g(x) = \sqrt{x}$ is a continues function for all $x > 0$, by Exercise 2.2.1 of [2], it follows that

$$\lim_{n \rightarrow \infty} \Pr \left(\left| \sqrt{\frac{\chi_n^2}{n}} - 1 \right| < \varepsilon \right) = 1 \quad \text{for every} \quad \varepsilon > 0. \tag{12}$$

Therefore, for all continuity points x of $F(x)$,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \tag{13}$$

where $F_n(x)$ is distribution function of random variable $\sqrt{\frac{\chi_n^2}{n}}$ and $F(x)$ is a degenerate distribution at 1. Using Lemma 1.5.6 of [17], the set $\{\alpha; \lim_{n \rightarrow \infty} F_n^{-1}(\alpha) \neq F^{-1}(\alpha)\}$ is at most countable

and hence a set of Lebesgue measure zero, where $G^{-1}(\alpha) = \inf_x \{x; G(x) \geq \alpha\}$ is the α -quantile of the distribution function G . Hence, Theorem 1.6.3 in [17] implies that

$$m_{[0,1]} \left(\left\{ \alpha; \lim_{n \rightarrow \infty} \sqrt{\frac{\chi_{n,\alpha}^2}{n}} \neq 1 \right\} \right) = 0, \quad \text{for all } 0 < \alpha < 1. \quad (14)$$

Obviously, we can conclude that $m_{[0,1]} \left(\left\{ \alpha; \lim_{n \rightarrow \infty} \sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}} \neq 1 \right\} \right) = 0$ and

$$m_{[0,1]} \left(\left\{ \alpha; \lim_{n \rightarrow \infty} \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \neq 1 \right\} \right) = 0. \quad \text{Therefore the set}$$

$$\left\{ \alpha; \lim_{n \rightarrow \infty} \left[\sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}}, \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \right] \neq [1, 1] \right\} \quad (15)$$

is of Lebesgue measure zero. Equivalently by Lemma 3, we may conclude

$$\begin{aligned} & \left\{ \alpha; \lim_{n \rightarrow \infty} \left[R \left(\hat{C}_p \right) \sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}}, R \left(\hat{C}_p \right) \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \right] \neq \left[R \left(\hat{C}_p \right), R \left(\hat{C}_p \right) \right] \right\} = \\ &= \left\{ \alpha; \lim_{n \rightarrow \infty} \left[R \left(\hat{C}_p \otimes \sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}} \right), R \left(\hat{C}_p \otimes \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \right) \right] \neq \left[R \left(\hat{C}_p \right), R \left(\hat{C}_p \right) \right] \right\} = \\ &= \left\{ \alpha; \lim_{n \rightarrow \infty} \left[\hat{C}_p \otimes \sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}}, \hat{C}_p \otimes \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \right] \neq \left[\hat{C}_p, \hat{C}_p \right] \right\}, \quad \text{by (9)} \end{aligned}$$

is a set of Lebesgue measure zero, and its compliment is a set of Lebesgue measure one. Hence (10) is proved.

Remark 1. The $100(1 - \alpha)\%$ fuzzy confidence interval for \tilde{C}_p , given in Theorem 1, shows a good reaction to the sample size n . In fact as n increases, the length of this interval decreases. In particular as n tends to infinity, the limit of confidence intervals length goes to the point estimate of \tilde{C}_p .

Remark 2. When the process specification limits $U(a_u, b_u, c_u)$ and $L(a_l, b_l, c_l)$ are precise numbers, then $a_u = b_u = c_u$ and $a_l = b_l = c_l$. In other words when specification limits are indicator functions, then as a result of Theorem 2, one can conclude that

$$m_{[0,1]} \left(\left\{ \alpha; \lim_{n \rightarrow \infty} \left[\hat{C}_p \sqrt{\frac{\chi_{n-1,\alpha/2}^2}{n-1}}, \hat{C}_p \sqrt{\frac{\chi_{n-1,1-\alpha/2}^2}{n-1}} \right] = \left[\hat{C}_p, \hat{C}_p \right] \right\} \right) = 1. \quad (16)$$

See Remark 5.4 of [11].

5. CONCLUSIONS

If we define the specification limits by fuzzy quantities, it is more appropriate to define the process capability indices as fuzzy numbers. Although we can obtain a point estimate for these fuzzy process capability indices, but we would never expect this point estimate to be exactly equal to the parameter value. Therefore, we usually compute a $100(1 - \alpha)\%$ confidence interval for our parameter. In this research, we proved an open problem related to the behavior of the $100(1 - \alpha)\%$ confidence interval for \tilde{C}_p stated in [11, 13]. Actually, we proved that as the sample size tends to infinity, the limit of this confidence interval's length goes to the point estimate of the capability index \tilde{C}_p .

Acknowledgements

The first and second authors are partially supported by grant No. 853143-1385-8-15 from National Science Foundation of Iran and Fuzzy Systems and its Applications Center of Excellence, Shahid Bahonar University of Kerman, Kerman, Iran.

REFERENCES

- [1] Chung, K. L., A Course in Probability Theory. Harcourt, Brace World, Inc, New York, 1968.
- [2] Davison, A. C., Statistical Models. Cambridge University Press, Cambridge, U.K., 2003.
- [3] Fortemps, P., Roubens, M., Ranking and defuzzification methods base on area compensation, Fuzzy Sets and Systems 82 (1996) 319-330.
- [4] Juran, J. M., Juran's Quality Control Handbook. 3rd edition, McGraw-Hill, New York, 1974.
- [5] Kane, V. E., Process capability indices. Journal of Quality Technology 18 (1986), pp. 41-52.
- [6] Kotz, S., Lovelace, C. R., Process Capability Indices in Theory and Practice. Oxford University Press Inc., New York, 1998.
- [7] Lee, H. T., index estimation using fuzzy numbers. European Journal of Operational Research 129 (2001), pp.683-688.
- [8] Maleki, H. R., Ranking functions and their applications to fuzzy linear programming. Far East Journal of Mathematical Sciences 4 (2002),pp.283-301.
- [9] Moeti, M. T., Parchami, A., Mashinchi, M., A note on fuzzy process capability indices, Iranica Scientia 13 (4, (2006), pp.379-385.
- [10] Nguyen, H. T., Walker, E. A., A First Course in Fuzzy Logic. Chapman & Hall/CRC, London, 3rd edition, 2006.
- [11] Parchami, A., Mashinchi, M., Maleki, H. R., Fuzzy confidence intervals for fuzzy process capability index. Journal of Intelligent & Fuzzy Systems 17 (3, (2006), pp. 287-295.
- [12] Parchami, A., Mashinchi, M., Fuzzy estimations for process capability indices. Information Sciences 177 (6), (2007),pp.1452-1462.
- [13] Parchami, A., Mashinchi, M., Maleki, H. R., Confidence interval for fuzzy process capability index. WSEAS Transactions on Systems 4 (5), (2005), pp. 546-551.
- [14] Parchami, A., Mashinchi, M., Yavari, A. R., Maleki, H. R., Process capability indices as fuzzy numbers, Austrian Journal of Statistics 34 (4), (2005), pp. 391-402.
- [15] Roubens, M., Inequality constraints between fuzzy numbers and their use in mathematical programming. Stochastic versus fuzzy approaches to multiobjective mathematical programming under uncertainty. Kluwer Academic Publishers, (1991), pp. 321-330.
- [16] Sadeghpour-Gildeh, B., Comparison of and process capability indices in the case of measurement error occurrence. Proceedings of IFSA World Congress, Istanbul, Turkey, (2003), pp. 563-567.
- [17] Serfling, R. J., Approximation Theorems of Mathematical Statistics. John Wiley and Sons, New York, 1980.
- [18] Yongting, C., Fuzzy quality and analysis on fuzzy probability. Fuzzy Sets and Systems 83, (1996), 283-290.



Abbas Parchami - He was born in 1977. He received his B.Sc. (2001) and M.Sc. (2004) both in statistics from Ferdowsi University of Mashhad and Shahid Bahonar University of Kerman (SBUK) in Iran, respectively. He is currently working as a lecturer at department of statistics of SBUK. His research interests include fuzzy statistics, decision making and quality control.



Mashaallah Mashinchi - He was born in 1951 in Kerman, Iran. He received his B.Sc. (1976) and M.Sc. (1978) both in statistics from Ferdowsi University of Mashhad, Shiraz University in Iran, respectively, and his Ph.D. (1987) in mathematics from Waseda University in Japan. He is a professor in the department of statistics at the Shahid Bahonar University of Kerman (SBUK) in Iran, Editor-in-Chief of the Iranian Journal of Fuzzy Systems and the director of Fuzzy Systems and its Applications Center of Excellence at SUBK. His current interest is in fuzzy mathematics, especially on statistics, decision making and algebraic systems.



Vahid Partovi Nia - He was born in 1980 and finished his B.Sc. in Ferdowsi University of Mashhad (FUM) in 2002. After achieving top ranks in national M.Sc. entrance competition in different branches of statistics, he chose Mathematical Statistics in FUM the Iran's Center of Excellence in Statistics at that time. Then he joined chair of statistics of Swiss Federal Institute of Technology in Lausanne as teaching assistant and started his Ph.D. on high-dimensional classification and clustering in 2005. He is interested in methodological problems motivated from applications.