The Likelihood Interpretation of Fuzzy Data

Marco Cattaneo

School of Mathematics and Physical Sciences
University of Hull

SMPS 2016, Rome, Italy
12 September 2016

introduction

- fuzzy set (Zadeh, 1965): described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a crisp set

introduction

- fuzzy set (Zadeh, 1965): described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a crisp set

- in order to use fuzzy sets to model the reality, an (operational) interpretation for the numerical values of μ is needed

introduction

- fuzzy set (Zadeh, 1965): described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a crisp set
- in order to use fuzzy sets to model the reality, an (operational) interpretation for the numerical values of μ is needed
- the interpretation does not need to be unique, and can be based on analogies or hypothetical situations

introduction

- fuzzy set (Zadeh, 1965): described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a crisp set

- in order to use fuzzy sets to model the reality, an (operational) interpretation for the numerical values of μ is needed

- the interpretation does not need to be unique, and can be based on analogies or hypothetical situations

- only the rules of fuzzy set theory implied by the considered interpretation should be used in applications

introduction

- fuzzy set (Zadeh, 1965): described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a crisp set
- in order to use fuzzy sets to model the reality, an (operational) interpretation for the numerical values of μ is needed
- the interpretation does not need to be unique, and can be based on analogies or hypothetical situations
- only the rules of fuzzy set theory implied by the considered interpretation should be used in applications
- e.g., probability has several interpretations (in part based on analogies or hypothetical situations), but they all imply the rules of probability theory (at least on finite spaces)

likelihood interpretation

- membership = likelihood: $\mu : \mathcal{X} \to [0,1]$ is the likelihood function *lik* on $\mathcal{X}$ induced by the observation of an event D:

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) \quad \text{for all} \quad x \in \mathcal{X}$$

likelihood interpretation

- membership = likelihood: $\mu : \mathcal{X} \to [0,1]$ is the likelihood function *lik* on $\mathcal{X}$ induced by the observation of an event D:

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) \quad \text{for all} \quad x \in \mathcal{X}$$

- e.g., if we get the information that "John is tall", we can model it by a fuzzy set with membership function $\mu : \mathcal{X} \to [0,1]$ with $\mu(x) \propto P(D \mid x)$, where the elements of $\mathcal{X}$ are the possible values of John's height in cm, and $P(D \mid x)$ is the probability of the event D of getting the information that "John is tall" when John's height is x cm

likelihood interpretation

- membership = likelihood: $\mu : \mathcal{X} \to [0,1]$ is the likelihood function *lik* on $\mathcal{X}$ induced by the observation of an event D:

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) \quad \text{for all} \quad x \in \mathcal{X}$$

- e.g., if we get the information that "John is tall", we can model it by a fuzzy set with membership function $\mu : \mathcal{X} \to [0,1]$ with $\mu(x) \propto P(D \mid x)$, where the elements of $\mathcal{X}$ are the possible values of John's height in cm, and $P(D \mid x)$ is the probability of the event D of getting the information that "John is tall" when John's height is x cm

- oldest interpretation of fuzzy sets (Black, 1937; Menger, 1951; Loginov, 1966; Hisdal, 1988; Dubois et al., 1997; Singpurwalla and Booker, 2004; Coletti and Scozzafava, 2004; Cattaneo, 2008): epistemic interpretation (μ describes some information about the value of $x \in \mathcal{X}$), whose exact meaning depends on the interpretation of probability

likelihood interpretation

- membership = likelihood: $\mu : \mathcal{X} \to [0,1]$ is the likelihood function *lik* on $\mathcal{X}$ induced by the observation of an event D:

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) \quad \text{for all} \quad x \in \mathcal{X}$$

- e.g., if we get the information that "John is tall", we can model it by a fuzzy set with membership function $\mu : \mathcal{X} \to [0,1]$ with $\mu(x) \propto P(D \mid x)$, where the elements of $\mathcal{X}$ are the possible values of John's height in cm, and $P(D \mid x)$ is the probability of the event D of getting the information that "John is tall" when John's height is x cm

- oldest interpretation of fuzzy sets (Black, 1937; Menger, 1951; Loginov, 1966; Hisdal, 1988; Dubois et al., 1997; Singpurwalla and Booker, 2004; Coletti and Scozzafava, 2004; Cattaneo, 2008): epistemic interpretation (μ describes some information about the value of $x \in \mathcal{X}$), whose exact meaning depends on the interpretation of probability

- often overlooked: likelihood defined only up to a multiplicative constant (because otherwise depending on irrelevant information), and likelihood interpretation thus restricted to normalized fuzzy sets (i.e., satisfying $\sup_{x \in \mathcal{X}} \mu(x) = 1$)

fuzzy data

- direct generalization to fuzzy data of all statistical methods based on the likelihood function (including all Bayesian methods): the likelihood function *lik* on Θ induced by the fuzzy observation $\mu(x) \propto P(D \mid x)$ is

$$lik(\theta \mid D) \propto P(D \mid \theta) \propto \underbrace{\int_{\mathcal{X}} \mu(x) \, dP(x \mid \theta)}_{\text{probability of fuzzy observation } \mu \text{ (Zadeh, 1968)}} \quad \text{for all} \quad \theta \in \Theta$$

fuzzy data

- direct generalization to fuzzy data of all statistical methods based on the likelihood function (including all Bayesian methods): the likelihood function *lik* on Θ induced by the fuzzy observation $\mu(x) \propto P(D \,|\, x)$ is

$$lik(\theta \,|\, D) \propto P(D \,|\, \theta) \propto \underbrace{\int_{\mathcal{X}} \mu(x)\, dP(x \,|\, \theta)}_{\text{probability of fuzzy observation } \mu \text{ (Zadeh, 1968)}} \quad \text{for all} \quad \theta \in \Theta$$

- conjunction of independent fuzzy sets μ_1, μ_2 (i.e., induced by events D_1, D_2 that are conditionally independent given x):

$$\mu(x) \propto P(D_1 \cap D_2 \,|\, x) \propto \underbrace{\mu_1(x)\, \mu_2(x)}_{\text{product rule (Zadeh, 1968)}} \quad \text{for all} \quad x \in \mathcal{X}$$

fuzzy data

- direct generalization to fuzzy data of all statistical methods based on the likelihood function (including all Bayesian methods): the likelihood function *lik* on Θ induced by the fuzzy observation $\mu(x) \propto P(D \mid x)$ is

$$lik(\theta \mid D) \propto P(D \mid \theta) \propto \underbrace{\int_{\mathcal{X}} \mu(x)\, dP(x \mid \theta)}_{\text{probability of fuzzy observation } \mu \text{ (Zadeh, 1968)}} \quad \text{for all} \quad \theta \in \Theta$$

- conjunction of independent fuzzy sets μ_1, μ_2 (i.e., induced by events D_1, D_2 that are conditionally independent given x):

$$\mu(x) \propto P(D_1 \cap D_2 \mid x) \propto \underbrace{\mu_1(x)\, \mu_2(x)}_{\text{product rule (Zadeh, 1968)}} \quad \text{for all} \quad x \in \mathcal{X}$$

- likelihood function induced by fuzzy i.i.d. observations $\mu_1, \ldots, \mu_n$:

$$lik\left(\theta \mid \bigcap_{i=1}^{n} D_i\right) \propto \underbrace{\prod_{i=1}^{n} \int_{\mathcal{X}} \mu_i(x)\, dP(x \mid \theta)}_{\text{often used (Gil and Casals, 1988; Denœux, 2011) on the basis of (Zadeh, 1968)}} \quad \text{for all} \quad \theta \in \Theta$$

fuzzy inference

- built-in fuzzy statistical inference: the fuzzy set $\mu = \mathit{lik}$ on Θ describes the information about θ obtained from the data

fuzzy inference

- built-in fuzzy statistical inference: the fuzzy set $\mu = lik$ on Θ describes the information about θ obtained from the data

- extension principle = profile likelihood: given a function $g : \Theta \to \Gamma$, the profile likelihood function μ_g on Γ is defined by

$$\mu_g(\gamma) = \sup\{\mu(\theta) : \theta \in \Theta,\ g(\theta) = \gamma\} \quad \text{for all} \quad \gamma \in \Gamma$$

fuzzy inference

- built-in fuzzy statistical inference: the fuzzy set $\mu = \mathit{lik}$ on Θ describes the information about θ obtained from the data

- extension principle = profile likelihood: given a function $g : \Theta \to \Gamma$, the profile likelihood function μ_g on Γ is defined by

$$\mu_g(\gamma) = \sup \{\mu(\theta) : \theta \in \Theta,\ g(\theta) = \gamma\} \quad \text{for all} \quad \gamma \in \Gamma$$

- α-cuts = likelihood-based confidence intervals: in agreement with the extension principle, the likelihood-based confidence interval for $\gamma = g(\theta)$ with cutoff point $\alpha \in (0, 1)$ is

$$\{\gamma \in \Gamma : \mu_g(\gamma) > \alpha\}$$

conclusion

- an interpretation of membership function values is needed in each real-world application, and only the rules of fuzzy set theory implied by the considered interpretation should be used

conclusion

- an interpretation of membership function values is needed in each real-world application, and only the rules of fuzzy set theory implied by the considered interpretation should be used

- the likelihood interpretation is very natural, but its connection with probability can be confusing: likelihood and probability are complementary descriptions of uncertainty

conclusion

- an interpretation of membership function values is needed in each real-world application, and only the rules of fuzzy set theory implied by the considered interpretation should be used
- the likelihood interpretation is very natural, but its connection with probability can be confusing: likelihood and probability are complementary descriptions of uncertainty
- with the likelihood interpretation, fuzzy data and fuzzy inferences are perfectly compatible with standard statistical analyses

conclusion

- an interpretation of membership function values is needed in each real-world application, and only the rules of fuzzy set theory implied by the considered interpretation should be used
- the likelihood interpretation is very natural, but its connection with probability can be confusing: likelihood and probability are complementary descriptions of uncertainty
- with the likelihood interpretation, fuzzy data and fuzzy inferences are perfectly compatible with standard statistical analyses
- future work will focus on the rules of fuzzy set theory implied by the likelihood interpretation (besides extension principle, product rule, and normalization)

references

Black, M. (1937). Vagueness. *Philos. Sci.* 4, 427–455.

Cattaneo, M. (2008). Fuzzy probabilities based on the likelihood function. In *Soft Methods for Handling Variability and Imprecision*. Springer, 43–50.

Coletti, G., and Scozzafava, R. (2004). Conditional probability, fuzzy sets, and possibility: a unifying view. *Fuzzy Sets Syst.* 144, 227–249.

Denœux, T. (2011). Maximum likelihood estimation from fuzzy data using the EM algorithm. *Fuzzy Sets Syst.* 183, 72–91.

Dubois, D., Moral, S., and Prade, H. (1997). A semantics for possibility theory based on likelihoods. *J. Math. Anal. Appl.* 205, 359–380.

Gil, M. Á., and Casals, M. R. (1988). An operative extension of the likelihood ratio test from fuzzy data. *Stat. Pap.* 29, 191–203.

Hisdal, E. (1988). Are grades of membership probabilities? *Fuzzy Sets Syst.* 25, 325–348.

Loginov, V. I. (1966). Probability treatment of Zadeh membership functions and their use in pattern recognition. *Eng. Cybern.* 4, 68–69.

Menger, K. (1951). Ensembles flous et fonctions aléatoires. *C. R. Acad. Sci.* 232, 2001–2003.

Singpurwalla, N. D., and Booker, J. M. (2004). Membership functions and probability measures of fuzzy sets. *J. Am. Stat. Assoc.* 99, 867–877.

Zadeh, L. A. (1965). Fuzzy sets. *Inf. Control* 8, 338–353.

Zadeh, L. A. (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* 23, 421–427.