Building possibility distribution based on confidence intervals of parameters of Gaussian mixtures

Mohammad Ghasemi Hamed, Mathieu Serrurier, Nicolas Durand

Building possibility distribution based on confidence intervals of parameters of Gaussian mixtures

M. Ghasemi Hamed [1,2], M. Serrurier[1], and N. Durand[1,2]

[1] IRIT - Université Paul Sabatier
118 route de Narbonne 31062, Toulouse Cedex 9, France
[2] DTI-R&D DGAC -
7 avenue Edouard Belin 31400 Toulouse, France

Abstract. In parametric methods, building a probability distribution from data requires an a priori knowledge about the shape of the distribution. Once the shape is known, we can estimate the optimal parameters value from the data set. However, there is always a gap between the estimated parameters from the sample sets and true parameters, and this gap depends on the number of observations. Even if an exact estimation of parameters values might not be performed, confidence intervals for these parameters can be built. One interpretation of the quantitative possibility theory is in terms of families of probabilities that are upper and lower bounded by the associated possibility and necessity measure. In this paper, we assume that the data follow a Gaussian distribution, or a mixture of Gaussian distributions. We propose to use confidence interval parameters (computed from a sample set of data) in order to build a possibility distribution that upper approximate the family of probability distributions whose parameters are in the confidence intervals. Starting from the case of a single Gaussian distribution, we extend our approach to the case of Gaussian mixture models.

1 Introduction

In 1978, Zadeh introduced the possibility theory [5] as an extension of his theory of fuzzy sets. Possibility theory offers an alternative to the probability theory when dealing with partial knowledge or epistemic uncertainty. A possibility distribution contains all the probability distributions that are respectively upper and lower bounded by the possibility and the necessity measure. In this scope, a probability-possibility distribution has been proposed [3]. Our method is based on the probability-possibility transformation [3] We assume that we have an a priori knowledge about the shape of the distribution, hence we have to estimate the unknown parameters of the distribution with respect to the data. Considering the amount of available data , it may be illusionary to expect to have an exact (or even a good) estimation of these parameters. In the case of Gaussian distributions, statistical approaches can be employed in order to build confidence intervals. Under some assumptions, the same kind of intervals can be

computed for Gaussian mixture. The uncertainty associated with the parameters value cannot be encoded by the estimated probability distribution. However, in some critical domain, such as risk management, taking into account this kind of uncertainty can be a crucial issue. The idea of the paper is to handle this type of uncertainty by building possibility distributions that bound the set of probability distributions, which have parameters in the confidence intervals. Thus, the possibility distribution built represents the family of acceptable probability distributions given a set of data and knowing the shape of the distribution. This paper is structured as follows: we begin with a background on the possibility theory and more specifically on the probabilistic interpretation of the possibility theory and the probability-possibility transformation. Then we see how to build the mean and variance confidence intervals for a Gaussian distribution with respect to a set of data. In the third section, we propose a method for constructing the maximal specific possibility distributions that bounds all the Gaussian distributions which have their parameters in their corresponding confidence intervals. Finally, we extend these results to Gaussian mixture models under some assumption and simplifications.

2 Background

A possibility distribution π is a function from Ω to $(\mathbf{R} \rightarrow [0,1])$. It has been proved in [2] that a possibility distribution π represents the family of the probability distributions Θ for which the measure of each subset of Ω will be bounded its possibility measures. In the following, we will note p the density function of a probability distribution (sometimes referred directly as probability distribution) and P its cumulative distribution function. Given a probability distribution p, a confidence interval I_α is a subset of Ω such as $P(I_\alpha) = \alpha$. A possibility measure Π is equivalent to the family Θ of probability measures such that $\Theta = \{P | \forall A \subseteq \Omega, P(A) \leq \Pi(A)\}$. We define I^*_α, also referred as quantile, as the smallest confidence interval with probability measure equal to α (this interval is unique only if p have finite number of modes). In many cases it is desirable to move from the probability framework to the possibility framework. Dubois et al.[3] suggest that when moving from the possibility to probability framework we should use the "maximum specificity" principle which aims at finding the most informative possibility distribution. Formally the maximum specificity principle is defined as follow. This kind of transformation (probability to possibility) may be desirable when we are in presence of weak source of knowledge or when it is computationally harder to work with the probability measure than with the possibility measure. The "most specific" possibility distribution function for a finite mode probability distribution function has the following formula [3] :

$$\pi_t(x) = sup\{1 - P(I^*_\alpha), x \in I^*_\alpha\} \tag{1}$$

where π_t is the "most specific" possibility distribution, I^*_α is the α confidence interval.

In the following, we will note $g(\mu, \sigma^2)$ for a Gaussian distribution of mean μ and variance σ^2, $g(x, \mu, \sigma^2)$ for its density function and $G(x, \mu, \sigma^2)$ for its cumulative distribution function. Having a set of n pieces of data, the confidence interval of probability 0.95 of the mean μ, evaluated from the data, is obtained as follows :

$$\mu - 1.96 * \frac{\sigma}{\sqrt{n}} < \mu < \mu + 1.96 * \frac{\sigma}{\sqrt{n}}. \tag{2}$$

The confidence interval of probability $\beta = 1 - \alpha$ of the variance σ^2, evaluated from the data, is obtained as follows :

$$\frac{n-1}{\chi^2_{1-\frac{\beta}{2},n-1}}\sigma^2 < \sigma^2 < \frac{n-1}{\chi^2_{\frac{\beta}{2},n-1}}\sigma^2 \tag{3}$$

where $\chi^2_{\frac{\beta}{2},n-1}$ is the density of Chi-squared distribution with $n-1$ degrees of freedom evaluated for $x = \frac{\beta}{2}$. In what follows, we will denote respectively confidence interval of the mean and the variance by $[\mu_{min}, \mu_{max}]$ and $[\sigma^2_{min}, \sigma^2_{max}]$, given a confidence level.

3 Possibility distribution for a family of Gaussian distribution

When data follows a normal distribution with unknown mean μ and standard deviation σ, a direct estimation of these parameters may be risky if only a low amount of data are available. In the previous section, we have described the confidence intervals for means and variance given a set of data (for simplification, we always take $\alpha = 0.95$ for the intervals in the following). If the estimation of these parameters is a critical issue of a decision process, it may be interesting to take into account the normal distributions that may have generated the data. In this scope, we propose to construct the most specific possibility distribution that contains the family $\Theta = \{g(\mu, \sigma^2) | \mu \in [\mu_{min}, \mu_{max}], \sigma^2 \in [\sigma^2_{min}, \sigma^2_{max}]\}$ of Gaussian distributions which have mean and variance parameters in the confidence intervals. We name Φ the set of possibility distributions obtained by transforming each distribution in Θ. So $\Phi = \{\pi | \pi = Tr(p), p \in \Theta\}$ where $Tr(g)$ is the probability-possibility transformation of a g.

Proposition 1 *Given $\Theta = \{g(\mu, \sigma^2) | \mu \in [\mu_{min}, \mu_{max}], \sigma^2 \in [\sigma^2_{min}, \sigma^2_{max}]\}$ the family of Gaussian distributions which have mean and variance parameters in the confidence intervals, the possibility distribution defined by*

$$\pi_\Theta(x) = Sup\{\pi(x), \pi \in \Phi\}$$

encodes all the probability family Θ. π_Θ has the following definition :

$$\pi_\Theta(x) = \begin{cases} 1 & \text{if } x \in [\mu_{min}, \mu_{max}] \\ 1 - 2 * G(x, \mu_{min}, \sigma^2_{max}) & \text{if } x < \mu_{min} \\ 1 - 2 * G(2 * \mu_{max} - x, \mu_{max}, \sigma^2_{max}) & \text{if } x > \mu_{max} \end{cases}$$

Where $[\mu_{min}, \mu_{max}]$ *is the mean confidence interval* $[\sigma^2_{min}, \sigma^2_{max}]$ *is the variance confidence interval.*

Proposition 2 *: The possibility distribution* π_Θ *is the most specific possibility distribution which encodes all the Gaussian distributions of the family* Θ*.*

In this section, we have described the construction of π_Θ and we have proved that π_Θ is the most specific distribution that encodes all the Gaussian distributions that have parameters inside the confidence intervals.

4 Possibility distribution for a family of Gaussian mixture model

In this section, we propose to build a possibility distribution that encodes a family of Gaussian Mixture Model (GMM). As for the simple Gaussian case, the idea is to compute the confidence interval of the parameters of an estimated from a data set. Since the possibility distribution that bounds all of these GMMs may be difficult to compute and not handy to use, we compute directly a trapezoid possibility distribution that describes faithfully this family. A Gaussian mixture model is a weighted sum of K different Gaussian distribution as described by the equation below:

$$p(x) = \sum_{i=1}^{k} \omega_i * g(x, \mu_i, \sigma_i^2) \tag{4}$$

where ω_i is the weight of the ith Gaussian density, μ_i its mean and σ_i^2 its variance. The mixture weights has to satisfy the following constraint : $\sum_{i=1}^{k} \omega_i = 1$. There are several techniques available for estimating the GMM parameters from data, Maximum Likelihood Estimation being the most popular one [4]. In the following, we assume that our sample set comes from an unknown GMM density function with a known number of components. We follow the same idea than in Section 3. We note Γ the family of GMM distributions which have parameters inside the confidence intervals. The computation of the trapezoid distribution for family of GMM is done by following these steps :

1. Computing the confidence intervals of the parameters and identifying the family of GMMs Γ.+
2. Computing the maximal possibility degree for the modes of the components.
3. Identifying the 0.05-cut of the possibility distribution which encodes the whole GMMs family.
4. Building the trapezoid distribution.

Confidence intervals of the parameters of a GMM. We only consider the confidence intervals for the parameters of the Gaussian component densities, and we let the weights ω_i constant. As pointed out previously, the rationale behind the definition of GMMs is that our observations comes from a set of independent Gaussian distributions. In this scope, we assume that the parameters μ_i and σ_i^2

of each Gaussian are estimated with respect to $n_i = \omega_i * n$ piece of data (where n is the size of the dataset used for building the GMM), i.e. the ratio of data equals to the weights associated to the distribution. Once we have n_i, the confidence interval of the parameters μ_i and σ_i^2 are computed as previously.

Identification of the modes. The modes of a GMM are the local maximum of the density function. However, it has been shown that the identification of the modes of a GMM is a hard problem [1]. In order to simplify the calculus, we will compute the maximum possibility degree for the means (i.e. the modes) of the Gaussian components. Then, we compute intervals by taking into account the mean confidence intervals. These intervals will be used in order to build the trapezoid distribution.

Identification of the 0.05**-cuts.** As for the family of Gaussian distributions, we want that the 0.05-cuts of the trapezoid $trap_\Gamma$ contains all the 95% quantiles of the GMMs in Γ. We obtain the values of these intervals by considering the two extreme largest GMM densities in Γ which is the followings :

$$\forall, i \; g_{extrem}(x) = \sum_{i}^{k} \omega_i * g(x, \mu_{imin}, \sigma^2_{jmax})(x)$$

Building the trapezoid distribution $trap_\Gamma$ Having the set of triples $m_i = \{\mu_{imin}, \mu_{imax}, \pi^*(\mu_i)\}$ and the interval $[a_{0.05}, d_{0.05}]$, we can define $trap_\Gamma = (a, b, c, d)$ where $[a, d]$ is the support of the distribution and $[b, c]$ its core. $trap_\Gamma$ is the convex hull of points defined by the triples, which bounds the possibility value of the modes of the GMMs, and the 0.05-cuts. Thus, the lower bound of the core b is the lowest intersection between the line $y = 1$ and the lines that cross the point $(\mu_{imin}, \pi^*(\mu_i))$ and the point $(a_{0.05}, 0.05)$.

The trapezoid $trap_\Gamma$ has the following properties :

- it upper-bounds all the probability-possibility transformation of the modes of all the components of the GMMs in Γ.
- the 0.05-cuts of $trap_\Gamma$ contains all the 95% quantiles of the GMMs in Γ.

However, it is not guaranteed that $trap_\Gamma$ upper-bound all the probability-possibility transformation the GMMs in Γ, even for the α-cuts with $\alpha \geq 0.05$.

5 Illustration

For illustration, we consider a set of data that are generated by the GMM $p(x) = 0.7 * g(x, \mu_1 = -2, \sigma_1 = 0.8) + 0.3 * g(x, \mu_2 = 4, \sigma_2 = 1.5)$. We suppose that the parameters have been estimated with 200 pieces of data. Figure 1 illustrates this results for $n = 200$ and $n = 50$.

We can observe that the size of the core and the support increase quickly when the number of data decreases. This is due to the fact that we assume that each component is estimated independently with the corresponding ratio of data. This makes that the more the GMM complex is (i.e. the more component it has) the larger the dataset is needed in order to have an acceptable estimation of the parameters.

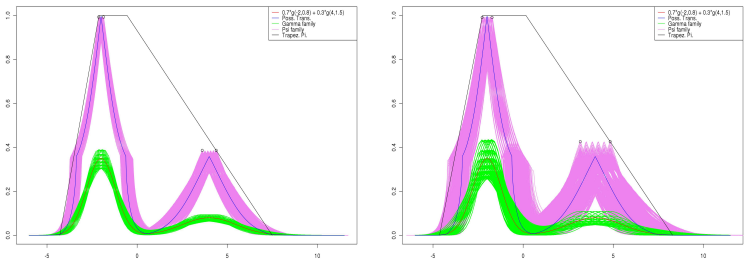

Fig. 1. $trap_{\Gamma}$ for the distribution $p(x) = 0.7 * g(x, -2, 0.8) + 0.3 * g(x, 4, 1.5)$ with $n = 200$ on the left and $n = 50$ on the right.

6 Conclusion

In this paper, we have described how to construct a possibility distribution from a set of data that is generated from a Gaussian distribution, or a Gaussian mixture model. We first compute confidence intervals for the parameters. Then, we built a possibility distribution that contains all the Gaussian distributions that have their parameters into their corresponding confidence intervals. In the case of GMM, we build a trapezoid that has good properties with respect to the modes and the 0.95 quantile. This approach can be useful in domains where a high level of confidence is required, due to safety or security reasons (aeronautic, medical applications, risk analysis). The method has the advantage to compute a possibility distribution that encodes both the probabilistic knowledge, the uncertainty due to the amount of data available and the complexity of the shape of the probability distribution. In the future, we will focus on the exact computation of the mode of the GMMs in order to have a better approximation of the most specific possibility distribution that encodes Γ. We also plan to embed this method into machine learning approaches such as Bayesian classifiers or regression when confidence intervals of the error are computed by local estimation.

References

1. Miguel Á. Carreira-perpiñán. Mode-finding for mixtures of gaussian distributions. Technical report, Dept. of Computer Science, University of Sheffield, 1999.
2. D. Didier. Possibility theory and statistical reasoning. *Computational Statistics and Data Analysis*, 51:47–69, 2006.
3. D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. *Reliable Computing*, 10:2004, 2004.
4. G. McLachlan. Mixture models. *Marcel Dekker, New York*, 1988.
5. L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100(Supplement 1):9–34, 1999.