Composition of Probability Density Functions Based on Minimization of Kullback-Leibler Divergence

J. Kracík

Department of Adaptive Systems, Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic, Prague, Czech Republic
e-mail: kracik@utia.cas.cz

I. Introduction

This paper deals with composition of probability density functions (pdfs) in a sense of composition of information pieces. As a natural example of such a composition we can give a living creature forming its image of the environment using information provided by its senses. Such information pieces are of different nature, they may be partially contradictory to each other, or some of them may be missing. In spite of that, living creatures are able to process such information completely naturally.

In this sense, roughly speaking, pdf composition means to find a joint pdf for given particular "marginal" pdfs. The given pdfs are arbitrary, and random quantities, for which the particular pdfs are given, may overlap. Therefore, a joint pdf, the marginal distributions of which are equal to the given pdfs, need not exist. What we are looking for is a joint pdf which is, in a certain sense, close to all given "marginal" pdfs. Such kind of composition is needed in applications like, e.g., sensor fusion [1] or multiple participant decision making [2].

II. Problem Formulation

For simplicity, in this paper we use a particular example instead of general formulation, nevertheless, a general task of this type can be formulated in the exactly same way and the outcomes presented in this paper are applicable as well.

Let us consider three random quantities x_1, x_2, x_3, two pdfs $f_1(x_1, x_2), f_2(x_2, x_3)$, and nonnegative coefficients α_1, α_2, such that $\alpha_1 + \alpha_2 = 1$. What we are searching is a joint pdf $f(x_1, x_2, x_3)$ whose marginal pdfs are, in a certain sense, close to the given pdfs $f_1(x_1, x_2), f_2(x_2, x_3)$ with respect to the weights α_1, α_2. To be close to this two pdfs are contradictory requirements. From this point of view, the marginals of the resulting pdf $f(x_1, x_2, x_3)$ provide a certain compromise between $f_1(x_1, x_2)$ and $f_2(x_2, x_3)$ and the coefficients α_1, α_2 determine "how close" the resulting pdf should be to the given ones – the higher α_1 is the closer to $f_1(x_1, x_2)$ the corresponding marginal pdf of $f(x_1, x_2, x_3)$ should be (and similarly for α_2). The coefficients α_1, α_2 can represent our belief to the corresponding sources of information.

A solution, which would be suitable from our point of view, should fulfil certain "natural" requirements:

- The given pdfs are supposed to be inconsistent in general, i.e., a joint pdf, say $g(x_1, x_2, x_3)$, such that $f_1(x_1, x_2)$ and $f_2(x_2, x_3)$ are marginal pdfs of $g(x_1, x_2, x_3)$ need not exist.
- The link between information represented by the given pdfs and the resulting one should be clear.
- The result of the composition should not depend on the order in which the given pdfs are processed.
- The process of composition should not add any information, which is not in the given pdfs.
- The considered variables may be discrete as well as continuous.
- Uniqueness of the resulting pdf is to be guaranteed, if it is necessary.

Of course, there are many publications (e.g., [3], [4]) dealing with similar problems, however, none of them fulfil all of the preceding requirements.

1. Minimization of a Weighted Sum of Kullback-Leibler Divergences

It can be shown that all of the requirements are fulfilled if we formulate the composition as an optimizing task – minimization of a weighted sum of Kullback-Leibler divergences of given pdfs and corresponding marginal pdfs of the "common approximation", i.e.,

$$f(x_1,x_2,x_3) \in \operatorname*{Argmin}_{\tilde{f}(x_1,x_2,x_3)} \alpha_1 D(f_1(x_1,x_2)||\tilde{f}(x_1,x_2)) + \alpha_2 D(f_2(x_2,x_3)||\tilde{f}(x_2,x_3)), \tag{1}$$

where $D(f(x)||g(x)) = \int f(x) \ln \frac{f(x)}{g(x)} \mathrm{d}x$ is a Kullback-Leibler divergence [5] of pdfs $f(x)$ and $g(x)$, $\tilde{f}(x_1,x_2), \tilde{f}(x_2,x_3)$ are marginal pdfs of $\tilde{f}(x_1,x_2,x_3)$.

It is proved [6] that $f(x_1,x_2,x_3)$ fulfils (1) iff it holds

$$\alpha_1 f_1(x_1,x_2) f(x_3|x_1,x_2) + \alpha_2 f_2(x_2,x_3) f(x_1|x_2,x_3) = f(x_1,x_2,x_3). \tag{2}$$

Unfortunately, an analytical solution of equations corresponding to (2) is known for a few, very simple, cases only. However, for discrete quantities an approximate solution can be found using an iterative algorithm described in [6]. The algorithm is based on repetitive use of an operator A, which assigns to an arbitrary pdf $f(x_1,x_2,x_3)$ the pdf

$$Af(x_1,x_2,x_3) = \alpha_1 f_1(x_1,x_2) f(x_3|x_1,x_2) + \alpha_2 f_2(x_2,x_3) f(x_1|x_2,x_3). \tag{3}$$

If we denote the weighted sum of KL divergences in (1) $\mathcal{D}(\tilde{f})$, i.e.,

$$\mathcal{D}(\tilde{f}) = \alpha_1 D(f_1(x_1,x_2)||\tilde{f}(x_1,x_2)) + \alpha_2 D(f_2(x_2,x_3)||\tilde{f}(x_2,x_3))$$

then for an arbitrary pdf $f(x_1,x_2,x_3)$ it holds

$$\mathcal{D}(Af) \leq \mathcal{D}(f)$$

with equality iff f is solution of (2). It can be shown that for an arbitrary initial approximation $f_0(x_1,x_2,x_3)$ the sequence $f_0,\ Af_0,\ A(Af_0),\dots$ converges to the solution of (2)[6].

Such an algorithm can be theoretically used for continuous quantities as well, nevertheless, in this case the approximations do not remain in any "reasonable class" of pdfs, for instance, finite mixtures of pdfs from a certain exponential family. For this reason, the approximations can be hardly represented in a computer, whereas for discrete quantities the approximations are easily represented by a finite grid.

2. Example

Using of the iterative algorithm based on (3) we demonstrate on a simple example for which an analytical solution is not known. Let us consider two discrete random quantities x_1, x_2 with values in $\{1,2,\dots,80\}$, three pdfs $f_1(x_1), f_2(x_2), f_3(x_1,x_2)$, and weights $\alpha_1,\alpha_2,\alpha_3$. The corresponding operator A has the following form

$$Af(x_1,x_2) = \alpha_1 f_1(x_1) f(x_2|x_1) + \alpha_2 f_2(x_2) f(x_1|x_2) + \alpha_3 f_3(x_1,x_2).$$

The pdfs $f_1(x_1), f_2(x_2), f_3(x_1,x_2)$ are selected as discretized Gaussian pdfs (Fig. 1), (Fig. 2), (Fig. 3) and the starting approximation $f_0(x_1,x_2)$ is a uniform pdf. Ratios of the weights $\alpha_1 : \alpha_2 : \alpha_3$ are 3:2:4. The approximation after 100 iterations $f(x_1,x_2)$ is in (Fig. 4). Evolution of $\mathcal{D}(f)$ during iterations is plotted in (Fig. 5).

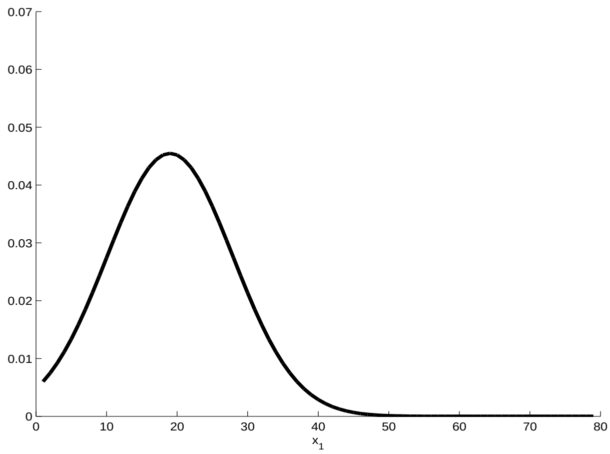

Figure 1: $f_1(x_1)$

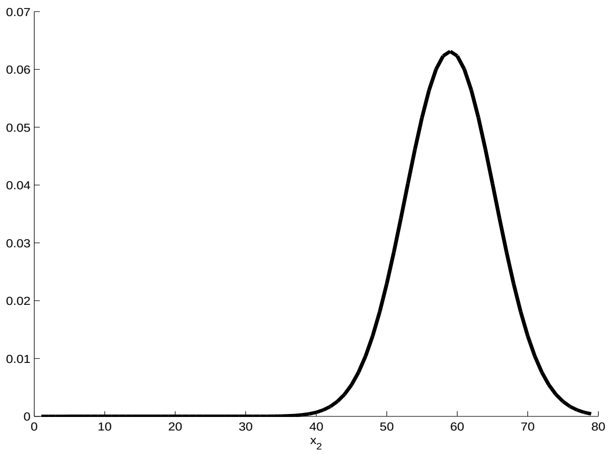

Figure 2: $f_2(x_2)$

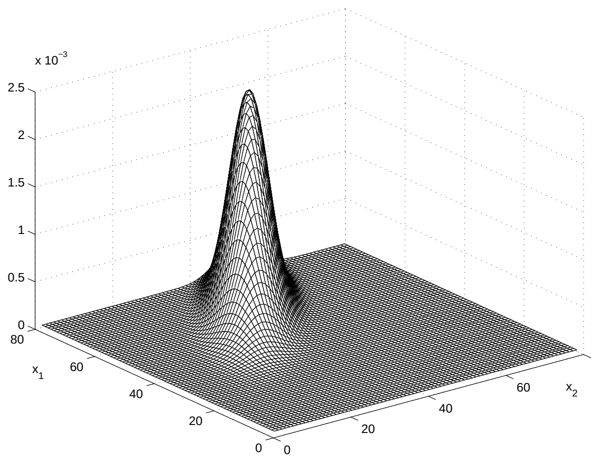

Figure 3: $f_3(x_1, x_2)$

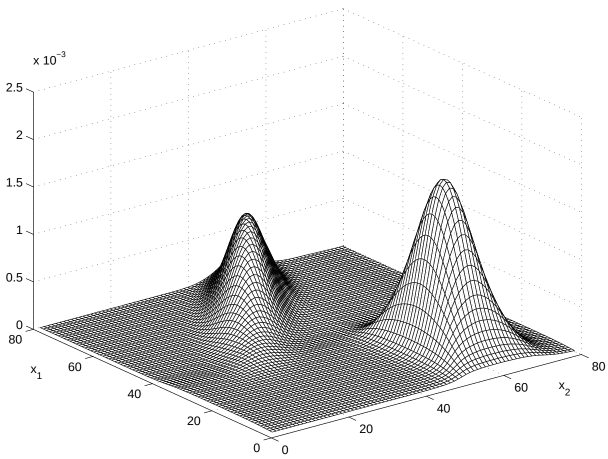

Figure 4: $f(x_1, x_2)$

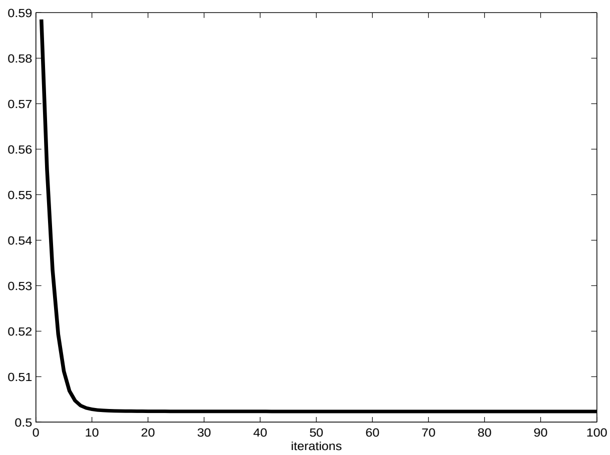

Figure 3: Evolution of $\mathcal{D}(f)$

III. Concluding Remarks

We have formulated composition of probability density functions as an optimizing task. The functional minimized (1) is a weighted sum of Kullback-Leibler divergences of given particular pdfs and corresponding marginal pdfs of their "common approximation". Composition of pdfs defined in such a way is suitable for applications in which a “global information” must be composed from, potentially inconsistent, “information pieces”.

An analytical solution of (1) is known only for a few special cases. In general, it must be solved approximately. An iterative algorithm for approximate solution is suggested in [6]. This algorithm is potentially able to find an arbitrarily good approximation, nevertheless, many problems remain to be solved to make it practically useful. The most important are:

- Find a practically usable modification of the algorithm for continuous quantities.
- Find an efficient stopping rule for the algorithm.

Acknowledgement

This work was supported by GAČR 102/03/0049, AV ČR S1075351, and AV ČR 1ET 100750401 BADDYR.

References

[1] M. Kárný, K. Warwick, T. Guy, J. Kracík, and I. Goodhew, “Framework for multisensory convergence,” in *Multiple Participant Decision Making*, J. Andrýsek, M. Kárný, and J. Kracík, Eds., vol. 9 of *International Series on Advanced Intelligence*, pp. 63–74. Advanced Knowledge International, Adelaide, Australia, 2004.

[2] M. Kárný and T. Guy, “On dynamic decision-making scenarios with multiple participants,” in *Multiple Participant Decision Making*, J. Andrýsek, M. Kárný, and J. Kracík, Eds., vol. 9 of *International Series on Advanced Intelligence*, pp. 17–28. Advanced Knowledge International, Adelaide, Australia, 2004.

[3] R. Jiroušek, “On experimental system for multidimensional model development MUDIN,” *Neural Network World*, (5):513–520, 2003.

[4] F. Jensen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York, 2001.

[5] S. Kullback and R. Leibler, “On information and sufficiency,” *Annals of Mathematical Statistics*, 22:79–87, 1951.

[6] J. Kracík, “Composition of probability density functions - optimizing approach,” Tech. Rep. 2099, ÚTIA AV ČR, Praha, 2004.