Weighted Probabilistic Opinion Pooling Based on Cross-Entropy

Vladimíra Sečkárová(✉)

Institute of Information Theory and Automation of the CAS,
Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic
seckarov@utia.cas.cz

Abstract. In this work we focus on opinion pooling in the finite group of sources introduced in [1]. This approach, heavily exploiting Kullback-Leibler divergence (also known as cross-entropy), allows us to combine sources' opinions given in probabilistic form, i.e. represented by the probability mass function (pmf). However, this approach assumes that sources are equally reliable with no preferences on, e.g., importance of a particular source. The discussion about the influence of the combination by preferences among sources (represented by weights) and numerical demonstration of the derived theory on an illustrative example form the core of this contribution.

Keywords: Minimum cross-entropy principle · Kullback-Leibler divergence · Linear opinion pooling · Combining probability distributions

1 Introduction

In this work we focus on decision making in the finite group (set) of sources, providing their opinions about the underlying (studied) problem to each other. The problem relates to a hidden (stochastic) phenomenon, that is not directly observable, but about which an opinion can be formulated. Examples of such phenomena are anticipated elections results, companies contracts and many others. The assumption on obtaining opinions yields a specific decision making process commonly known as opinion pooling. Due to the complexity of the space of possible decisions we consider the probability distributions over this set rather than single values. The final decision (result of pooling) is then a combination of probability distributions provided by sources.

Combining probabilistic information within a group of sources has been of interest for a long time, see, e.g., [2]. Different approaches consider different levels of cooperation among sources and different exploitations of their proposed combination. One can assume the sources are represented as a group of individuals "who must act together as a team and reach consensus" [3]. Or, one can consider that sources are "perfectly coherent, rational as decision makers and cooperate in agreeing to adopt a group utility function" [4].

Many probability combining approaches exploit information theory, namely, information divergences such as the Kullback-Leibler (KL) divergence [5] (the

S. Arik et al. (Eds.): ICONIP 2015, Part II, LNCS 9490, pp. 623–629, 2015.
DOI: 10.1007/978-3-319-26535-3_71

term cross-entropy is also used). Based on the order of its arguments we arrive at two basic classes - linear pools and log-linear pools, see, e.g., [6].

Linear pools consider, e.g., in [7], where "expert opinion is represented as a probability and associated with a confidence level that expresses the conviction of the corresponding expert on its own judgement".

The examples of KL-divergence based log-linear pools include, e.g., estimation [8], or determination of weights for prior distributions pooling [9].

Here, we follow the idea of combination derived under certain assumptions on cooperation between sources [10]. In particular, we focus on approach recently introduced in [1], which heavily exploits cross-entropy and yields a linear combination of discrete probability distribution. The formula for this combination was derived under the *wise-selfish* cooperation scenario assumption. *Wise* source is willing to cooperate and share its information with other sources in the group, but *selfishly* requires the result of combining to be "close" in the sense of bounded KL-divergence. There were no assumptions on preferences among sources reflecting their reliability or importance.

In this contribution we study how the values of this opinion pool changes when preferences about sources reliability or importance are known prior to combining. First, we briefly summarize the combining introduced in [1]. Then, we discuss the results when weights are included. To demonstrate the proposed idea we give an illustrative example.

2 Opinion Pooling of Discrete Probability Distributions Based on Cross-Entropy

As considered in [1], let us have a finite number of sources, $j = 1, \ldots, s < \infty$. Assume also that each source provides its opinion in the probabilistic form: as a probability mass function (pmf) assigning probability to each of $n < \infty$ outcomes of stochastic phenomenon:

$$\boldsymbol{p}_j = (p_{j1}, \ldots, p_{jn}) : \quad p_{ji} > 0, \ \sum_{i=1}^{n} p_{ji} = 1, \quad j = 1, \ldots, s. \tag{1}$$

Let $\boldsymbol{q}$ represent the combination of $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$. To obtain $\hat{\boldsymbol{q}}$, the estimator of $\boldsymbol{q}$, we search for the minimizer of the expected loss [11]. In particular, we search for the minimizer of the conditional expected value of the KL-divergence (KLD) with respect to the conditional probability density function (pdf) $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$ conditioned on $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$

$$\mathrm{E}_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)} \mathrm{KLD}(\boldsymbol{q}||\hat{\boldsymbol{q}}) = \mathrm{E}_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)} \sum_{i=1}^{n} q_i \ln \frac{q_i}{\hat{q}_i}. \tag{2}$$

The minimizer is the conditional expected value of $\boldsymbol{q}$ with respect to $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$

$$\hat{\boldsymbol{q}} = \mathrm{E}_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)}[\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s]. \tag{3}$$

The estimator (3) heavily depends on the form of the unknown conditional pdf $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$. To obtain this pdf we introduce the notion of the source's *selfishness* represented by the constraints on the expected KL-divergences from $\boldsymbol{p}_j$ to $\boldsymbol{q}$ with respect to the conditional pdf $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$. Moreover, we assume the sources consider the following equalities among the expected values of the KL-divergence:

$$\mathrm{E}_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)}[\mathrm{KLD}(\boldsymbol{p}_s||\boldsymbol{q})|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s] = \mathrm{E}_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)}[\mathrm{KLD}(\boldsymbol{p}_j||\boldsymbol{q})|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s], \tag{4}$$

$j = 1, \ldots, s-1$.

We exploit the minimum cross-entropy principle [12] and choose the conditional pdf $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$ that solves:

$$\min_{\pi(\boldsymbol{q}|\boldsymbol{p}_1,\ldots,\boldsymbol{p}_s)} \mathrm{KLD}(\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)||\pi_0(\boldsymbol{q})), \tag{5}$$

where $\pi_0(\boldsymbol{q})$ is the prior guess on the conditional pdf $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$.

The prior pdf is chosen as the pdf of the Dirichlet distribution for its computationally advantageous properties. The KL-divergence in (5) and the constraints in (4) can be then rewritten as follows

$$\ln \frac{\Gamma(\sum_{i=1}^n \nu_{0i})}{\Gamma(\sum_{i=1}^n \nu_{0i})} + \sum_{i=1}^n \ln \frac{\Gamma(\nu_{0i})}{\Gamma(\nu_i)} + \sum_{k=1}^n (\nu_k - \nu_{0k}) \left(\psi(\nu_k) - \psi(\sum_{l=1}^n \nu_l) \right) \tag{6}$$

and

$$-H(\boldsymbol{p}_j) + H(\boldsymbol{p}_s) = \sum_{i=1}^n (p_{ji} - p_{si}) \left(\psi(\nu_i) - \psi\left(\sum_{i=1}^n \nu_i \right) \right), \quad j = 1, \ldots, s-1, \tag{7}$$

where $\Gamma(.)$ is the Gamma function, $\psi(.)$ is the digamma function [13] and $\nu_{01}, \ldots, \nu_{0n}$ is the prior guess on parameters of the Dirichlet distribution. It can be shown that the conditional pdf $\pi(\boldsymbol{q}|\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)$ minimizing (5) is also the pdf of the Dirichlet distribution. Thus, it is satisfactory to perform the minimization in the set of possible values of parameters $\nu_1, \ldots, \nu_n$ instead of searching in the set of all possible probability distributions. The values of parameters, for which the value (5) is minimal, are denoted by $\hat{\nu}_1, \ldots, \hat{\nu}_n$.

The values of the parameters $\hat{\nu}_1, \ldots, \hat{\nu}_n$ are theoretically expressed by the following formula:

$$\hat{\nu}_i = \nu_{0i} + \sum_{j=1}^s \lambda_j (p_{ji} - p_{si}), \quad i = 1, \ldots, n, \tag{8}$$

where λ_j are the Lagrange multipliers resulting from minimization of (5) with respect to $(s-1)$ equations in (4).

The final weighted combination of $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$, represented by $\hat{\boldsymbol{q}}$ in (3), has then the following form:

$$\hat{q}_i = \frac{\hat{\nu}_i}{\sum_{k=1}^n \hat{\nu}_k} = \frac{\nu_{0i}}{\sum_{k=1}^n \nu_{0k}} + \frac{\sum_{j=1}^s \lambda_j (p_{ji} - p_{si})}{\sum_{k=1}^n \nu_{0k}}, \quad i = 1, \ldots, n, \tag{9}$$

since, based on (8), it holds that $\sum_{k=1}^{n} \hat{\nu}_k = \sum_{k=1}^{n} \nu_{0k}$.

To perform the numerical part of the optimization we exploited Matlab.

3 Cross-Entropy Based Opinion Pooling Influenced by Preferences Among Sources

The opinion pooling of sources' $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$ described above assumed no preferences among sources. If the preferences, represented by the weights $w_1, \ldots, w_s$ about sources $1, \ldots, s$, are available, they can influence the prior information about $(\nu_{01}, \ldots, \nu_{0n})$ and selfishness restrictions in (4). Both cases are discussed below.

For the guess on values of $(\nu_{01}, \ldots, \nu_{0n})$ it is optional to exploit the information known prior to combining. Since we focus on combining opinions about hidden (stochastic) phenomena with unavailable prior knowledge, we suggest to use the weighted arithmetic mean of $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$ as the prior guess:

$$\nu_{0i} = \sum_{j=1}^{s} w_j p_{ji}. \tag{10}$$

Since

$$\sum_{i=1}^{n} \nu_{0i} = \sum_{i=1}^{n} \sum_{j=1}^{s} w_j p_{ji} = \sum_{j=1}^{s} w_j, \tag{11}$$

we obtain that the part of combination (9) representing the part induced by the prior $\pi_0(\boldsymbol{q})$ is

$$p_{0i} = \frac{\nu_{0i}}{\sum_{k=1}^{n} \nu_{0k}} = \sum_{j=1}^{s} \frac{w_j}{\sum_{l=1}^{s} w_l} p_{ji}. \tag{12}$$

It is somewhat surprising that the Eq. (9) combines simultaneously both, the parameters of the Dirichlet distribution and pmfs $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s$. Provided pmfs can be viewed as individual guess for $(\nu_{01}, \ldots, \nu_{0n})$ when $\sum_{i=1}^{n} \nu_{0i} = 1$ (yielding $\sum_{j=1}^{s} w_j = 1$).

Besides the prior values $\nu_{01}, \ldots, \nu_{0n}$, it is also desirable that the constraints (4) will be affected by these weights. In particular, we approach the combination to more important sources by requiring

$$w_j \mathrm{E}\left[\mathrm{KLD}(\boldsymbol{p}_j || \boldsymbol{q}) | \boldsymbol{p}_1, \ldots, \boldsymbol{p}_s\right] = w_s \mathrm{E}\left[\mathrm{KLD}(\boldsymbol{p}_s || \boldsymbol{q}) | \boldsymbol{p}_1, \ldots, \boldsymbol{p}_s\right], \tag{13}$$

$j = 1, \ldots, s-1$, yielding the following weighted counterpart of constraints (7):

$$-w_j H(\boldsymbol{p}_j) + w_s H(\boldsymbol{p}_s) = \sum_{i=1}^{n} (p_{ji} w_j - p_{si} w_s) \left[\psi(\nu_i) - \psi\left(\sum_{k=1}^{n} \nu_k\right)\right]. \tag{14}$$

Then, the final combination $\hat{\boldsymbol{q}}$ in (9) is

$$\hat{q}_i = p_{0i} + \sum_{j=1}^{s-1} \frac{\lambda_j}{\sum_{l=1}^{s} w_l} (w_j p_{ji} - w_s p_{si}), \quad i = 1, \ldots, n. \tag{15}$$

with p_{0i} given in (12).

In the presented discussion, we have assumed that the elements of pmf provided by source j, $j = 1, \ldots, s$, have the same weights $w_{ji} = const$, $i = 1, \ldots, n$. If needed, also an element-dependent version w_{ji}, $i = 1, \ldots, n$ can be developed.

Let us now consider the following two sets of weights

$$\sum_{j=1}^{s} w_j = d \quad \text{and} \quad \sum_{j=1}^{s} w_j^* = \sum_{j=1}^{s} k w_j = d^*. \tag{16}$$

According to (10), the prior guesses on the parameters of the Dirichlet distribution look as follows:

$$\begin{aligned} &\text{for } w_1, \ldots, w_s: \quad \nu_{0i} = \textstyle\sum_{j=1}^{s} w_j p_{ji}, \\ &\text{for } w_1^*, \ldots, w_s^*: \quad \nu_{0i}^* = \textstyle\sum_{j=1}^{s} w_j^* p_{ji}. \end{aligned}$$

From the combination $\hat{\boldsymbol{q}}$ derived for weights w^*

$$\hat{q}_i = \sum_{j=1}^{s} \frac{k w_j p_{ji}}{\sum_{l=1}^{s} k w_l} + \sum_{j=1}^{s-1} \frac{\lambda_j}{\sum_{l=1}^{s} k w_l} (k w_j p_{ji} - k w_s p_{si}), \tag{17}$$

where $k = \frac{d^*}{d}$, it might seem that both optimal estimators, based on w and w^*, coincide. Recall the KL-divergence in (6) leading to $\hat{\boldsymbol{q}}$ by minimization with respect to $\nu_1, \ldots, \nu_n$. Due to non-linearity of this function in prior guess $\nu_{01}, \ldots, \nu_{0n}$, we can not expect the combinations based on $w_1, \ldots, w_s$ and $w_1^* = k w_1, \ldots, w_s^* = k w_s$ to be equal.

In the following illustrative example we demonstrate the change in values of $\hat{\boldsymbol{q}}$ for different set of weights $w_1, \ldots, w_s$.

4 Illustrative Example

Let us have 2 sources ($s = 2$) which provided the following pmfs ($n = 3$)

$$\begin{aligned} \boldsymbol{p}_1 &= [0.75, 0.05, 0.2], \\ \boldsymbol{p}_2 &= [0.3, 0.1, 0.6]. \end{aligned}$$

Consider the following weights:

- first set of weights: arithmetic mean (equal weights summing to one),
- second set of weights: first source fixed $w_1 = 0.1$, weight of the second source is rising,
- third set of weights: weights of both sources are rising ($w_2 = 2 \times w_1$).

Values of the combination $\hat{\boldsymbol{q}}$ of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ in the last instant, see Fig. 1, are given in Table 1.

In Fig. 1 we see that the resulting combination stabilizes quickly and with rising values of the weights (second set and third set of weights) the combination of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ based on (15) tends towards the arithmetic mean of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. This might be a consequence of the considered *selfish* scenario – the equality in the constraints in (14).

Table 1. Values of the combination $\hat{\boldsymbol{q}} = (\hat{q}_1, \hat{q}_2, \hat{q}_3)$ for different sets of weights (value given for the last considered value of weights).

Arithmetic mean:	(0.525, 0.075, 0.4)
First set of weights:	(0.403, 0.089, 0.508)
Second set of weights $(w_1 = 0.1,\ w_2 = 27.2)$:	(0.519, 0.076, 0.405)
Third set of weights $(w_1 = 13.6,\ w_2 = 27.2)$:	(0.522, 0.075, 0.403)

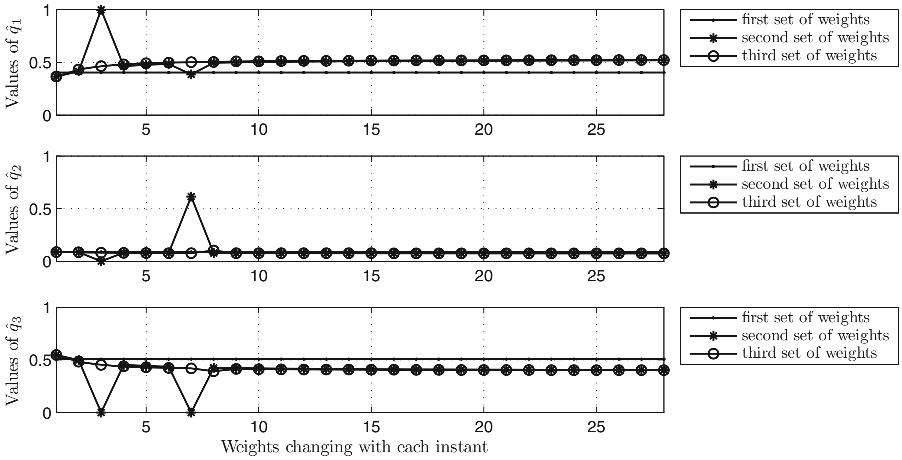

Fig. 1. Resulting combination of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ for different sets of weights. *First set of weights*: equal weights summing to one. *Second set of weights*: fixed weight of the first source ($w_1 = 0.1$), weight of the second source rising. *Third set of weights*: $w_2 = 2 \times w_1$, both weights rising.

5 Conclusion and Future Work

We focused on combining opinions provided in the probabilistic form based on the cross-entropy introduced in [1]. We discussed theoretically and numerically the change in the value of the combination when preferences among sources, represented by the weights, were included. The results of the illustrative example showed that, with rising values of the weights, the considered combining approach is stable and tends towards the arithmetic mean. Thus, the properties of suggested weighted version of the combination have to be properly investigated.

The further research includes development of another way how to involve preferences among sources, e.g., by allowing inequality in constraints (14).

Acknowledgments. This research has been supported by GAČR 13-13502S.

References

1. Sečkárová, V.: Minimum cross-entropy based weights in dynamic diffusion estimation in exponential family. Pliska Studia Math. Bulg. **24**, 181–188 (2015)
2. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition). Princeton Classic Editions. Princeton University Press, Princeton (2007)
3. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**(345), 118–121 (1974)
4. West, M.: Bayesian aggregation. J. R. Stat. Soc. (Ser. A) **147**, 600–607 (1984)
5. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
6. Abbas, A.E.: A Kullback-Leibler view of linear and log-linear pools. Decis. Anal. **6**(1), 25–37 (2009)
7. Garcia, M., Puig, D.: Robust aggregation of expert opinions based on conflict analysis and resolution. In: Conejo, R., Urretavizcaya, M., Pérez-de-la-Cruz, J.-L. (eds.) CAEPIA/TTIA 2003. LNCS (LNAI), vol. 3040, pp. 488–497. Springer, Heidelberg (2004)
8. Dedecius, K., Sečkárová, V.: Dynamic diffusion estimation in exponential family models. IEEE Sig. Process. Lett. **20**(11), 1114–1117 (2013)
9. Rufo, M.J., Martín, J., Pérez, C.J.: Log-linear pool to combine prior distributions: a suggestion for a calibration-based approach. Bayesian Anal. **7**(2), 411–438 (2012)
10. Kárný, M., Guy, T.V., Bodini, A., Ruggeri, F.: Cooperation via sharing of probabilistic information. Int. J. Comput. Intell. Stud. **1**(5), 139–162 (2009)
11. Bernardo, J.M.: Expected information as expected utility. Ann. Stat. **7**, 686–690 (1979)
12. Shore, J.E., Johnson, R.W.: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans. Inf. Theory **26**, 26–37 (1980)
13. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Applied Mathematics Series. Dover Publications, New York (1964)