Proceedings of Machine Learning Research 58 (2016) 41-50

Performance of Kullback-Leibler Based Expert Opinion Pooling for Unlikely Events

Vladimíra Sečkárová SECKAROV@UTIA.CAS.CZ
Department of Adaptive Systems
Institute of Information Theory and Automation
Pod Vodárenskou věží 4, 182 00 Prague 8, CZ

Editor: Tatiana V. Guy, Miroslav Kárný, David Rios-Insua, David H. Wolpert

Abstract

The aggregation of available information is of great importance in many branches of economics, social sciences. Often, we can only rely on experts' opinions, i.e. probabilities assigned to possible events. To deal with opinions in probabilistic form, we focus on the Kullback-Leibler (KL) divergence based pools: linear, logarithmic and KL-pool (Sečkárová, 2015). Since occurrence of events is subject to random influences of the real world, it is important to address events assigned lower probabilities (unlikely events). This is done by choosing pooling with a higher entropy than standard linear or logarithmic options, i.e. the KL-pool. We show how well the mentioned pools perform on real data using absolute error, KL-divergence and quadratic reward. In cases favoring events assigned higher probabilities, the KL-pool performs similarly to the linear pool and outperforms the logarithmic pool. When unlikely events occur, the KL-pool outperforms both pools, which makes it a reasonable way of pooling.

Keywords: Opinion Pooling, Combining Probability Distributions, Minimum Kullback-Leibler Divergence

1. Introduction

Problem of information aggregation from multiple sources is of interest in decision making and its applications in areas such as social sciences, economics and business. The choice of the final decision is a delicate process and if an unexpected situation occurs, it can have big psychological or financial impact. In this contribution, we address this problem by considering statistical and information-theoretic techniques and show that, if necessary, impact of the final decision in unexpected situations can be softened.

In applications of decision making (elections, information markets) we often rely on information formulated as an opinion. The process of aggregating this type of information is then referred to as (expert) opinion pooling. By the term expert we not only mean a person, who has knowledge about the variable of interest based on previous experience. We also include sources which, by investing reasonable time and energy, exploit the today's amount of information distributed by media into their advantage. Some of the experts may be certain about the decision (event), some prefer to express the uncertainty among possible events. To include the uncertainty in predicting possible events, we consider the expert opinion in the form of probability distribution. We assume, that experts are able to form probability distribution as a probability vector and we will not include the details of its elicitation; an extensive discussion about elicitation of probability distributions can be found in (Garthwaite et al., 2005).

Although sources form their opinion to their best knowledge and abilities and assign high probability to the most probably event, the occurrence of the described event is subject to the random influences of the real world such as nature, political situation, ... Then, an event assigned lower probability (an unlikely event) causing non-negligible loss may occur. Thus, the pool admitting unlikely events should assign more uncertainty in events; but still give a reasonable result in the regular case, when event assigned high probability from experts occurred. To measure the amount of uncertainty for probability distribution resulting from combining experts' opinions we exploit information theory, i.e. the Shannon entropy (Shannon and Weaver, 1963). To measure the utility of resulting probability distribution we consider the Kullback-Leibler (KL) divergence as recommended in (Bernardo, 1979).

There are many algorithms for combining probability distributions available and being developed in areas such as risk analysis (Clemen and Winkler, 1999), weather forecasting (Ranjan and Gneiting, 2010), economics (Chen et al., 2005), parameter estimation with knowledge elicitation (Kárný et al., 2014), knowledge sharing with deliberator (Azizi and Quinn, 2016). We focus on the basic combining algorithms used for opinion pooling: linear and logarithmic pool. Both can be obtained via unconstrained optimisation, i.e., minimisation of the KL-divergence (Abbas, 2009). We also consider more sophisticated version of linear pool (KLp) inspired by (Kárný et al., 2009) and introduced in (Sečkárová, 2015), arising from the constrained minimisation of the KL-divergence. The result of any pooling can be viewed as a compromise among considered sources; the compromise is usually derived to satisfy group aims and every included individual has to sacrifice its own aims. Consequently, when majority of experts' opinions are similar, the influence of an expert with different (opposite) opinion is suppressed.

We address these shortcomings by focusing on a combining approach KLp that yields combination with higher entropy than, e.g., linear or logarithmic pool. In particular, we consider constraints on acceptance of the resulting combination of opinions by individual sources, again in the sense of the KL-divergence. KLp, being a conservative compromise, is thus an appropriate pool when unlikely event occurs, especially if reward for formulated opinion is included.

The aim of this contribution is to verify our hypothesis, that the KLp behaves similarly to standard pooling options (linear, logarithmic) in regular case, but outperforms these in case of unlikely events. For comparison purposes, we use the absolute error and the KL-divergence with respect to the perfect prediction (probability 0 for the losing team) together with the quadratic reward.

Next section contains the overview of the construction of the KL-divergence pool KLp and shows behavior of the linear, logarithmic and KLp pools in terms of the entropy of pooled probabilities. In the third section we apply considered opinion pools on the real data obtained from contest for National Football League in USA (NFL games). The fourth section concludes the work.

2. Combining Experts' Discrete Distributions

Consider a finite number of experts labeled by $j = 1, \ldots, s$ providing discrete probability distributions represented by n-dimensional probability vectors:

$$\mathbf{p}_j = (p_{j1}, \ldots, p_{jn}) : \quad p_{ji} \geq 0, \ \sum_{i=1}^{n} p_{ji} = 1, \ n < \infty, \ j = 1, \ldots, s.$$

where n denotes the number of possible events, i.e., outcomes of an underlying random variable.

Let $\mathbf{q}$ denote an unknown probability vector representing the combination of $\mathbf{p}_1, \ldots, \mathbf{p}_s$. Its estimator $\hat{\mathbf{q}}$ is chosen as the minimiser the expected Kullback-Leibler divergence (Bernardo, 1979):

$$\hat{\mathbf{q}} \in \underset{\tilde{\mathbf{q}}(\mathbf{p}_1,\ldots,\mathbf{p}_s)}{\arg\min} \, \mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)} \mathrm{KLD}(\mathbf{q}||\tilde{\mathbf{q}}). \tag{1}$$

The minimisation task (1) follows the theory of Bayesian decision making (Savage, 1972) and yields:

$$\hat{\mathbf{q}} = (\hat{q}_1, \ldots, \hat{q}_n) = \mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s]. \tag{2}$$

The expectation in (2) depends on the conditional probability density function (pdf) $\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)$, which we specify in two consequent steps:

1. We assume that each expert, if considered as the 'best' representation of value of unknown $\mathbf{q}$, has a finite expected divergence from $\mathbf{q}$

   $$\mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[\mathrm{KLD}(\mathbf{p}_j||\mathbf{q})|\mathbf{p}_1, \ldots, \mathbf{p}_s] < \infty, \quad j = 1, \ldots, s, \tag{3}$$

   and he is willing to include other experts' opinions in the final combination if they represent $\mathbf{q}$ equally or worse in terms of (3). Applying this condition to every expert in the group we obtain that conditional pdfs have to satisfy:

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

   $j = 1, \ldots, s-1$. Previous attempts in terms of the bounded Kerridge inaccuracy (Kerridge, 1961) can be found in (Sečkárová, 2013).

2. We assume that the set of all pdfs satisfying (4) is non-empty. We exploit the minimum cross-entropy principle (Shore and Johnson, 1980) and choose the pdf solving the following optimisation task

   $$\min_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s) \text{ satisfying (4)}} \mathrm{KLD}(\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)||\pi_0(\mathbf{q})), \tag{5}$$

   with

   $$\pi_0(\mathbf{q}) = \pi_0(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)$$

   being the prior guess on the conditional pdf $\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)$.

The constrained optimisation task (5) yields

$$\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s) \propto \pi_0(\mathbf{q}) \prod_{i=1}^{n} q_i^{\sum_{j=1}^{s-1} \lambda_j (p_{ji} - p_{si})}, \tag{6}$$

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

To obtain more specific form of the combination (2) we next specify prior pdf in (5). This pdf is defined over $(n-1)$-dimensional probability simplex - a set of all $\mathbf{p}_j$. Based also on the relation given in (6), we exploit numerically appealing Dirichlet distribution. We then obtain that pdf $\pi(\mathbf{q}|\mathbf{p}_1, \ldots, \mathbf{p}_s)$ is also the pdf of the Dirichlet distribution. Its parameters have the following form:

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

where $\nu_{01}, \ldots, \nu_{0n}$ are parameters of the prior Dirichlet distribution, $\lambda_1, \ldots, \lambda_{s-1}$ are the Lagrange multipliers from the constrained optimisation task.

Based on the properties of the Dirichlet distribution

$$\mathrm{E}_{\pi(\mathbf{q}|\mathbf{p}_1,\ldots,\mathbf{p}_s)}[q_i|\mathbf{p}_1, \ldots, \mathbf{p}_s] = \frac{\hat{\nu}_i}{\sum_{i=1}^{n} \hat{\nu}_i},$$

the estimator (2) is

$$\hat{q}_i = \frac{\nu_{0i}}{\sum_{k=1}^{n} \nu_{0k}} + \sum_{j=1}^{s-1} \frac{\lambda_j}{\sum_{k=1}^{n} \nu_{0k}} (p_{ji} - p_{si}), \quad i = 1, \ldots, n. \tag{7}$$

Equation (7) represents the KL-pool (KLp) of expert opinions $\mathbf{p}_1, \ldots, \mathbf{p}_s$.

In many real-life inspired cases (elections, betting predictions), there is very little or no prior information available before processing expert opinions. We are thus forced to exploit given opinions in prior guess for parameters of the Dirichlet distribution. Since no new information is included in combining, it is natural that the sum of prior parameters ν_{0i} and parameters after combining $\hat{\nu}_i$ to be equal (see (2)). In particular, we set this sum of parameters to be equal to the number of opinions (or generally observations). Each parameter ν_{0i} is then assigned value relative to the arithmetic mean of the given opinions

$$\nu_{0i} = s \sum_{j=1}^{s} \frac{p_{ji}}{s} = \sum_{j=1}^{s} p_{ji}, \quad i = 1, \ldots, n. \tag{8}$$

For comparison purposes, we use

- linear pool (linp):

$$\hat{\mathbf{q}} \propto \sum_{j=1}^{s} w_j p_j, \tag{9}$$

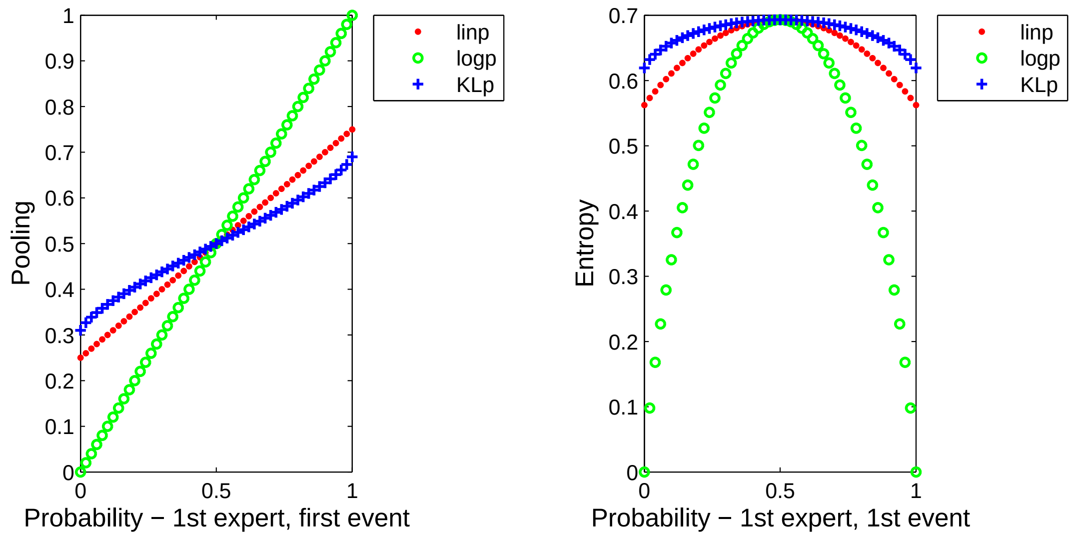

Figure 1: Values of pooling for $\hat{q}_1$ and amount of uncertainty (entropy) for the linear (9), logarithmic (10) and KL-pool (7).

- logarithmic pool (logp):

$$\hat{\mathbf{q}} \propto \prod_{j=1}^{s} p_j^{w_j}, \tag{10}$$

in this contribution, with w_j denoting weights, i.e., subjective preferences among experts $1, \ldots, s$. These two combining approaches arise from the minimisation of both versions of asymmetric KL-divergence without constraints (Abbas, 2009).

2.1 KL-pool as Treatment for Unlikely Events

We expect the experts posses certain level of expertise when assigning the probability to possible events. However, they often have incomplete information about random influences in real world such as nature, political situation, physical and mental abilities of others. Even minor changes in these factors can have huge impact on the occurrence of improbable events. Thus, it is important to reflect also events with lower probabilities assigned by experts – by allowing more uncertainty in the set of probable events. To measure the amount of uncertainty allowed by the combination in (7) we focus on the entropy (Rényi, 1961), exploited in environmental scenarios, e.g., methane emissions from wetlands (Sabolová et al., 2015). Higher values of entropy indicate that more uncertainty is present and that unlikely events should be assigned higher probability.

The formula (7) is not closed; the values of $\hat{\mathbf{q}} = (\hat{q}_1, \ldots, \hat{q}_n)$ are obtained numerically via optimisation with respect to either $\lambda_1, \ldots, \lambda_{s-1}$ or $\nu_1, ..., \nu_n$. The direct theoretical comparison of KLp with linp and logp, by exploiting constraints (4) rewritten for the Dirichlet distribution or by using an approximation of $\hat{\mathbf{q}}$, is a part of the future research.

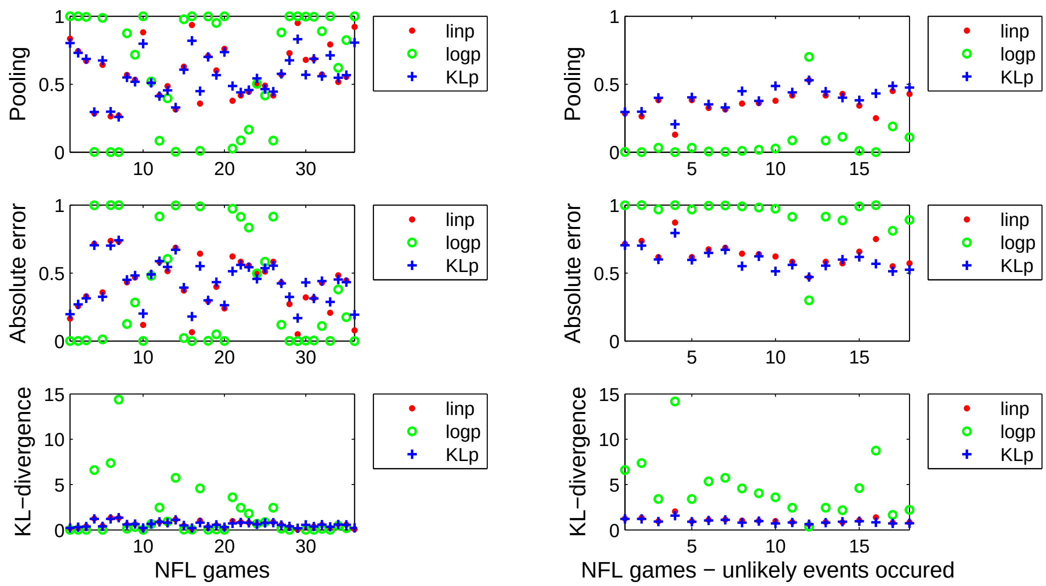

Figure 2: **Left column** Results for pooling based on the linear (linp), logarithmic (logp) and KL-pool (KLp), the absolute error (11) of the pools and the KL-divergence (13) for randomly picked games. **Right column** Results for pooling based on the linp, logp and KLp, the absolute error of the pools and the KL-divergence for the subset of games to which the players assigned low probabilities to the winning team.

We now consider a numerical example to see that the proposed KL-pool has this desirable property: let us thus combine two opinions $\mathbf{p}_j = (p_{j1}, p_{j2})$, $j = 1, 2$, where

$$\mathbf{p}_1 = (0, 1), \ldots, (1, 0) \quad \text{and} \quad \mathbf{p}_2 = (0.5, 0.5),$$

i.e., $\mathbf{p}_1$ varies in the above range with increments in p_{11} equal to 0.02. Figure 1 shows how KLp, with $\nu_{01}, \ldots, \nu_{0n}$ given by (8), behaves in comparison with linear (9) and logarithmic (10) pools arising for unit weights w_j, $j = 1, 2$. In Figure 1 on the left we see the outcomes of considered pools for the first event ($i = 1$). On the right, we see that the KL-pool has equal or higher entropy than linear and logarithmic pool for low-probability events ($p_{11} \leq 0.5$) and thus fits better for the treatment of unlikely events.

3. Real Data Application

In this section we demonstrate that the behaviour of the KL-pool, illustrated on the toy example in the previous section, is favourable when processing real data. We obtained data from `http://probabilityfootball.com/2005/`, an online football contest also referred to as “a game of skill”, where the players “estimate, for each football game, the probability that each NFL team will win based on how strong they believe each team to be”. Players pick a probability of winning for each team, ranging from 0 to 1 (according to webpage in percentile ranging from 0% to 100%), which they think accurately describes the strength of teams.

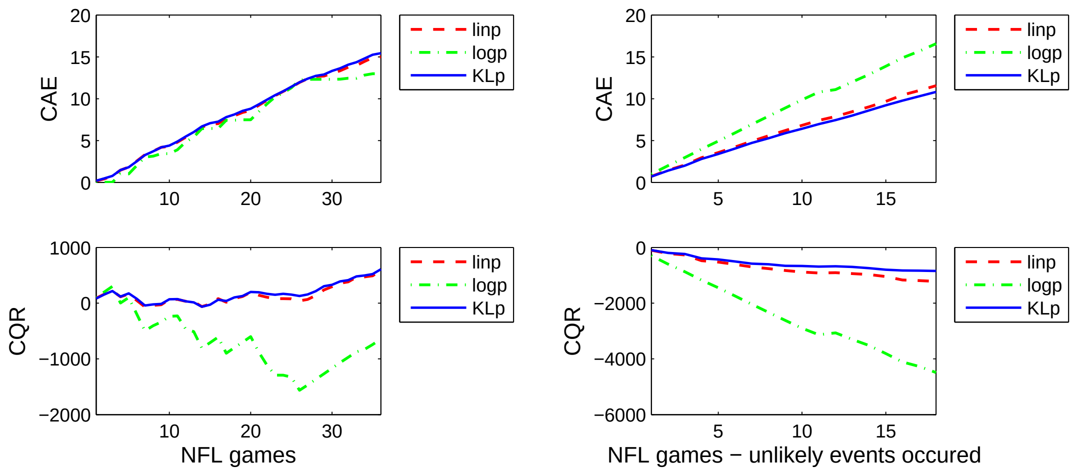

Figure 3: **Left column** Results for the linear (linp), logarithmic (logp) and KL-pool (KLp) in terms of cumulative absolute error (12) and cumulative quadratic reward (14) for randomly chosen games. **Right column** Results for the linp, logp and KLp in terms of CAE and CQR for games, when players assigned low probabilities to the winning team.

We apply previously mentioned linp (9), logp (10) and KLp (7) to demonstrate their behaviour and compare:

- values of pooling for the first event (the team that eventually won),
- absolute error measuring the difference between perfect prediction (probability 0 for losing team) and the actual prediction (Chen et al., 2005)

$$\mathrm{AE} = |0 - \hat{q}_{\text{losing team}}|, \quad \hat{q}_{\text{losing team}} \in \{\text{linp, logp, KLp}\} \tag{11}$$

  and its cumulative version

$$\mathrm{CAE}_t = \mathrm{CAE}_{t-1} + |0 - \hat{q}_{\text{losing team},t}|, \quad \hat{q}_{\text{losing team},t} \in \{\text{linp, logp, KLp}\} \tag{12}$$

  where CAE_{t-1} is the sum of absolute errors from $(t-1)$ previous games and $\hat{q}_{\text{losing team},t}$ is the pooling of opinions for current game,

- and KL-divergence

$$\mathrm{KLDiv} = \mathrm{KLD}(\mathbf{q}_{\text{perfect}}||\hat{\mathbf{q}}), \quad \hat{\mathbf{q}} \in \{\text{linp, logp, KLp}\} \tag{13}$$

  where $\mathbf{q}_{\text{perfect}}$ is a perfect prediction (probability 0 for losing team),

- cumulative quadratic reward

$$\mathrm{CQR} = \mathrm{CQR}_{t-1} + 100 - 400(\hat{q}_{\text{losing team},t})^2, \quad \hat{q}_{\text{losing team},t} \in \{\text{linp, logp, KLp}\} \tag{14}$$

  with quadratic reward used in the contest (Chen et al., 2005).

Firstly, focus on the first 6 weeks of NFL in 2005 for 7 successful players, each week 6 games were chosen. Results in the left column of Figure 2 show, that KL-pool performs similarly to the linear pool in terms of absolute error (11) and KL-divergence (13). This result is expected because of the additive nature of the KL-pool. The performance of the logarithmic pool is due to its multiplicative form oscillatory - it performs either very good or very poor.

Secondly, we focus on 18 games, when the majority of picked players put more believe into a team, which eventually lost the game. Figure 2 (in the right column) shows that in such case our method performs better than the linear pool, as intuitively expected.

Because of the design of this example, i.e., the variability in playing football teams, it is difficult to study the properties of considered pools sequentially (based on estimates from the previous game). Thus, we next exploit performance of these pools with respect to the cumulative versions of the absolute error (12) and the quadratic reward (14). The results depicted in Figure 3 are similar to the non-cumulative case: for randomly chosen games, the KL-pool performs similarly to the linear pool and both pools outperform the logarithmic pool (their CAE is lower and CQR is higher). In case of games, where the losing team was assigned lower probability by the players, the KL-pool outperforms both, the linear and the logarithmic pool.

4. Conclusion and Future Work

In this contribution we focused on problem of pooling expert opinions when events assigned lower probabilities occur (unlikely events). This is especially important in cases when also a reward for formulated opinion is also included: the higher probability assigned to the event that did not occur can yield huge loss. To treat this we need an opinion pool resulting in a combination with a higher entropy than standard pooling ways, so that unlikely events obtain reasonable probability. We considered the Kullback-Leibler based opinion pool (Sečkárová, 2015), values of which were obtained via constrained non-linear optimisation. This pool was constructed as the compromise for group of experts a) without sacrificing their own aims and b) without suppressing opinion which significantly differs from other opinions.

We showed on numerical example that KL-pool reaches higher values of entropy than generally known pools: (equally weighted) linear pool and logarithmic pool. We then applied these pools on real data and compared them using following performance measures: the absolute error, the KL-divergence and the quadratic reward. Because of the additive form of the KL-pool, its performance was similar to the performance of the linear pool for regular data (not many unlikely events occurred). In case of unlikely events, the KL-pool outperformed the linear and the logarithmic pool in terms of cumulative version of the absolute error and the quadratic reward and thus is a reasonable tool for pooling expert opinions. The future work includes the theoretical comparison of KL-pool with other opinion pools by using its approximation. Also, the application of the KL-pool on other sets of real data including betting with knowledge of fixed-odds, handling financial contracts, weather forecasts, is of interest.

Acknowledgements

This work has been supported by the GAČR grant GA16-09848S.

References

A. E. Abbas. A Kullback-Leibler View of Linear and Log-Linear Pools. *Decision Analysis*, 6(1):25–37, 2009.

S. Azizi and A. Quinn. Hierarchical fully probabilistic design for deliberator-based merging in multiple participant systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, PP(99):1–9, 2016. ISSN 2168-2216. doi: 10.1109/TSMC.2016.2608662.

J. M. Bernardo. Expected information as expected utility. *Ann. Stat.*, 7:686–690, 1979.

Y. Chen, C. Chu, T. Mullen, and D. M. Pennock. Information Markets vs. Opinion Pools: An Empirical Comparison. In *Proceedings of the 6th ACM Conference on Electronic Commerce*, pages 58–67, 2005.

R. T. Clemen and R. L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):1539–6924, 1999.

P. H. Garthwaite, J. B. Kadane, and A. O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701, 2005.

M. Kárný, T. V. Guy, A. Bodini, and F. Ruggeri. Cooperation via sharing of probabilistic information. *Int. J. of Computational Intelligence Studies*, 1(5):139–162, 2009.

M. Kárný, A. Bodini, T.V. Guy, J. Kracík, P. Nedoma, and F. Ruggeri. Fully probabilistic knowledge expression and incorporation. *Statistics and Its Interface*, 7(4):503–515, 2014.

D. F. Kerridge. Inaccuracy and inference. *J. R. Stat. Soc., Ser. B*, 23:184–194, 1961.

R. Ranjan and T. Gneiting. Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):71–91, 2010.

A. Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561, Berkeley, Calif., 1961. University of California Press.

R. Sabolová, V. Sečkárová, J. Dušek, and M. Stehlík. Entropy based statistical inference for methane emissions released from wetland. *Chemometrics and Intelligent Laboratory Systems*, 141(0):125 – 133, 2015. ISSN 0169-7439.

L. J. Savage. *The Foundations of Statistics.* Dover Books on Mathematics Series. Dover Publications, 1972.

V. Sečkárová. A Note on Weighted Combination Methods for Probability Estimation. In *Proccedings of the 3rd International Workshop on Scalable Decision Making held in conjunction with ECML/PKDD 2013 , Eds: Guy Tatiana V., Kárný Miroslav.* UTIA AV CR, Prague, 2013. ISBN 978-80-903834-8-7.

V. Sečkárová. Weighted probabilistic opinion pooling based on cross-entropy. In *Neural Information Processing, 22nd International Conference, ICONIP 2015, Istanbul, Turkey, November 9-12, 2015, Proceedings, Part II*, pages 623–629, Istanbul, Turkey, 2015.

C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Number pt. 11 in Illini books. University of Illinois Press, 1963.

J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory*, 26:26–37, 1980.