



Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

Václav Šmídl

Sequential Sampling Techniques for Log-Normal Combination of Probability Densities

No. 2244

December 12, 2008

1 Introduction

Bayesian approach to decision making under uncertainty relies on the fact that all probability densities involved in the problem are known. This assumption is justified in the case single of single decision-maker, however it is hard to meet in the case of distributed decision-making. In distributed scenario [4], the decision-makers exchange probability densities and use them within their own decision-making procedure. A methodology how to make use of such information was proposed in [5] using combination of probability densities.

1.1 Combination of probability densities

The field of combination of probability densities, also known as merging, is very rich. A survey in [3] provides commentary to 92 papers on the topic. From the range of available methods we select the following candidates:

- classical methods of arithmetic and geometric composition [3],
- Bayesian merging of probabilities based on log-normal ‘hyper’-model proposed in [5],

Methods not considered in this report typically require specification of arbitrary functions (e.g. ‘performance functions’ of [10]) or additional strategies of choosing weight in various variants of weighted arithmetic or geometric composition [3, 6].

An alternative to the chosen methods is was presented e.g. in [13] where an additional model of probability interaction is required. This model has a good Bayesian interpretation and can be used to adjust the merging procedure in a principled way. However, in this study, we will focus on the class of methods without such model.

1.2 Importance Sampling

Importance sampling [2] refers to a range of techniques for generating an empirical approximation of $f(x)$:

$$f(x) \approx f_\delta(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - x^{(i)}), \quad (1)$$

where $x^{(i)}$, $i = 1, \dots, n$ are independent identically distributed samples from the density $f(x)$ and $\delta(\cdot)$ denotes the Dirac δ -function. We reserve the symbol $f_\delta(\cdot)$ for the empirical density. Therefore, this approach is feasible only if the we can sample from the exact posterior, $f(x)$. If this is not the case, we can draw samples from a chosen proposal density (importance function), $q(x)$, as follows:

$$f(x) \approx \frac{f(x)}{q(x)} \frac{1}{n} \sum_{i=1}^n \delta(x - x^{(i)}). \quad (2)$$

Using the sifting property of the Dirac δ -function, the approximation can be written in the form of a *weighted* empirical density, as follows:

$$f_\delta(x) \approx \sum_{i=1}^n w_i \delta(x - x^{(i)}), \quad (3)$$

$$w_i \propto \frac{f(x)}{q(x)}. \quad (4)$$

Under this *importance sampling* procedure, the approximated density need only be evaluated point-wise. Furthermore, normalizing constant of $f(\cdot)$ is not required, since (3) can be normalized trivially via a constant $c = \sum_{i=1}^n w_i$.

The weighted empirical density can be converted into a non-weighted empirical density via a procedure known as *re-sampling* [2]. In principle, this procedure randomly creates a number of copies of each sample which is proportional the sample weight in expected value.

2 Merging of densities defined on the same multivariate variable

For simplicity, we will consider merging of finite number of densities, $f_s(x)$, $s = 1, \dots, S$, defined on variable x . The task is to merge all densities into one, $\tilde{f}(x)$.

2.1 Classical approaches

Arithmetic merging

$$\tilde{f}(x) = \frac{1}{S} \sum_{s=1}^S f_s(x). \quad (5)$$

Weighted variants of this approach were also proposed [3, 7].

Geometric merging

$$\tilde{f}(x) = \prod_{s=1}^S f_s(x)^{\frac{1}{S}}. \quad (6)$$

Weighted variants of this approach were also proposed [3].

2.2 Supra-Bayesian merging with log-normal model

It can be shown that both classical approaches can be obtained point-wise as expected values of Bayesian estimation for different models, [5]. In this approach, known as supra-Bayesian, merging of densities in a point \hat{x} is interpreted as parameter estimation, values of densities in \hat{x} are considered as data $d_s \equiv f_s(\hat{x})$ generated from a model $f(d|\theta)$, where $\theta = f(\hat{x})$ is the true underlying density. The unknown underlying density is estimated as posterior expectation

$$\tilde{f}(\hat{x}) \equiv \mathbf{E}_{f(d|d_1, \dots, d_S)}(d). \quad (7)$$

Arithmetic merging arise when d_s are normally-distributed with unknown mean μ and variance σ :

$$f(d|\mu, \sigma) = \mathcal{N}(\mu, \sigma).$$

Then, the posterior estimate

$$\mathbf{E}_{f(d|d_1, \dots, d_S)}(d) = \mathbf{E}_{f(\mu, \sigma|d)}(\mu) = \frac{1}{S} \sum_{s=1}^S d_s. \quad (8)$$

Substituting $d_s = f_s(x)$ into (8) yields (5).

Geometric merging arise as a special case ($\beta \rightarrow \infty$, see Appendix A) of the log-normal ‘hyper’-model [5]

$$f(\log(d)|\mu, \sigma) = \mathcal{N}(\mu, \sigma),$$

with prior on parameters μ, σ

$$f(\mu, \sigma) \propto \exp(-\beta\sigma), \quad \beta > 0. \quad (9)$$

Here, $\mu = f(x)$ is the value of the underlying density.

Analytical solution of the marginal

$$\mathbf{E}_{f(d|d_1, \dots, d_S)}(d) = \mathbf{E}_{f(\mu, \sigma|d)}(\exp(\mu + \frac{1}{2}\sigma)), \quad (10)$$

is a complicated function of BesselK functions, see Appendix A. Its general form can be greatly simplified for known value of S (i.e. the number of sources):

$$\mathbb{E}_{f(d|d_1, d_2)}(d) = \exp \left\{ \hat{\mu} + \sqrt{2\beta\lambda} \left(1 - 0.5 \sqrt{\frac{4\beta - 3}{\beta}} \right) \right\}, \quad (11)$$

$$\mathbb{E}_{f(d|d_1, d_2, d_3)}(d) = \frac{BesselK \left(0, \sqrt{2\beta\lambda} \sqrt{\frac{3\beta - 2}{3\beta}} \right)}{BesselK \left(0, \sqrt{2\beta\lambda} \right)} \exp(\hat{\mu}). \quad (12)$$

Here, $BesselK$ denotes modified Bessel function of the second kind, $\hat{\mu} = \frac{1}{S} \sum \ln(d_s)$, $S = S$, β is a parameter of prior density on σ , see Appendix B for discussion of its choice, and λ is a remainder after least squares,

$$\lambda = \sum \ln^2(d_s) - S\hat{\mu}^2.$$

Similar formulae to (11)–(12) can be derived for higher values of S . For even values of S , the result is of the form of (11) while results similar to (12) are obtained for odd values of S .

However, for practical purposes when S is higher or varying [5] proposed to use an approximate version of (10)

$$\mathbb{E}_{f(d|d_1, \dots, d_S)}(d) = \mathbb{E}_{f(\mu|d, \sigma)\delta(\sigma=\hat{\sigma})}(\exp(\mu + \frac{1}{2}\sigma)) = \exp \left\{ \hat{\mu} + \frac{-S^2 + (S+1)\sqrt{S^2 - 2S + 1 + 8\beta\lambda + 1}}{8\beta S} \right\}. \quad (13)$$

Remark 1. In effect, the only free parameters is the prior parameter β . It is shown in Appendix B that extreme values of this parameter, i.e. 1 and ∞ correspond to approximately-arithmetic and geometric merging, respectively. For practical purposes, a choice of β within this interval seems to be more reasonable option.

2.3 Merging of fragmental densities

The approach described in the previous Section can be applied only to densities defined on the same variable. Merging of incompatible sources—e.g. $f(x_1|x_2)$ and $f(x_1)$ —the approach can be used when the sources are extended—e.g. by multiplying by $f(x_2)$ and $f(x_2|x_1)$, respectively—to the common variables. In practical applications, it may be advantageous to seek principled extension for the particular approximation. In this paper, we assume that such extension is not available and thus we wish to minimize its impact on the resulting merger. An optimizing approach for this task was proposed in [5].

Consider a set of sources

$$F : \{f_s(x_{s,d}|x_{s,c})\}_{s=1}^S$$

where $x_{s,d}$ denotes multivariate variable on which the s th source is defined, and $x_{s,c}$ variable in condition of the s th source. The common variables is then a union of all variables

$$\mathbf{x} = \bigcup_{s=1}^S \{x_{s,d}, x_{s,c}\}.$$

We seek a global merger $\tilde{f}(\mathbf{x})$ which can be projected into a density on any combination of variables using standard probability calculus (i.e. marginalization and conditioning). It can be shown that optimum of $\tilde{f}(\mathbf{x})$ minimizing the impact of the extension to the merger is given by an implicit equation.

Algorithm 2. *The following iterative merging is solving the implicit equation [5] by successive approximations:*

1. Build initial guess of the merger on common variable $\tilde{f}^{(0)}(\mathbf{x})$, set iterative counter $j = 0$.
2. For each source,

(a) factorize $\tilde{f}^{(j)}(\mathbf{x})$ such that

$$\tilde{f}^{(j)}(\mathbf{x}) = \tilde{f}^{(j)}(x_{s,0}|x_{s,d}, x_{s,c})\tilde{f}^{(j)}(x_{s,d}|x_{s,c})\tilde{f}^{(j)}(x_{s,c}). \quad (14)$$

Here, $x_{s,0}$ denotes variables from \mathbf{x} that are not present in s th source.

(b) Create s th extended source by complementing the original source by projections of the current merger to the missing variables:

$$\tilde{f}_s^{(j)}(\mathbf{x}) = \tilde{f}^{(j)}(x_{s,0}|x_{s,d}, x_{s,c}) f_s(x_{s,d}|x_{s,c}) \tilde{f}^{(j)}(x_{s,c}). \quad (15)$$

3. Merge extended sources $\{\tilde{f}_i^{(j)}(\mathbf{x})\}$ into $\tilde{f}^{(j+1)}(\mathbf{x})$, using merging of densities on common variables.
4. If a divergence of $\tilde{f}^{(j)}(\mathbf{x})$ from $\tilde{f}^{(j+1)}(\mathbf{x}) > \text{threshold}$, set $j = j + 1$, goto 2. Otherwise, stop.

Note that the supra-Bayesian approach to probability merging requires to evaluate the likelihoods of the source pdfs including the normalizing constant. This is practically acceptable for merging of densities on the same variable, however, it poses a problem within the iterative procedure above where the resulting merger is must be re-normalized for the next iteration. Renormalization of the result is not a problem for the arithmetic mean (5) which can be easily re-normalized, neither for the geometric mean (6) which is also simple for exponential family of densities. However, renormalization of the log-normal merging is non-trivial and must be done numerically.

Remark 3 (Grid-based evaluation). Point-wise evaluation of the merger on a rectangular grid is appealing for evaluation of the normalization constant, marginal and conditional densities. In that case, the posterior density is approximated by a piecewise-linear density. However, well-positioned and fine grained grid may be build only for densities with limited support. Otherwise, the number of grid points may be too large to handle, especially for higher dimensional problems.

3 Importance sampling for probabilistic merging

Key properties of the log-normal merging—i.e. point-wise construction with subsequent renormalization—are consistent with assumptions of the importance sampling procedure.

3.1 Merging of weighted empirical densities

Consider a task of merging two densities (3) with common samples $x^{(i)}$ but different weights w_i . Merging is then performed at samples $x^{(i)}$, yielding the result in the form of weighted empirical density at the same points as the sources, but with modified weights, \tilde{w}_i .

This can be formalized as projection of the merger into weighted empirical density. Since we can not draw samples from the merger, we need to find a suitable proposal density. In the first step of algorithm 2, we may use the chosen initial density $\tilde{f}^{(0)}(\mathbf{x})$ as such proposal. At each sample, the non-normalized value of the merger can be computed at each point $\mathbf{x}^{(i)}$ using (10) where each source is computed using (15). The resulting merger after the first iteration is:

$$\tilde{f}^{(1)}(\mathbf{x}) \approx \sum_{i=1}^n \frac{\tilde{f}^{(1)}(\mathbf{x}^{(i)})}{f^{(0)}(\mathbf{x}^{(i)})} \delta(\mathbf{x} - \mathbf{x}^{(i)}) = \sum_{i=1}^n w_i \delta(\mathbf{x} - \mathbf{x}^{(i)}). \quad (16)$$

$$w_i = \frac{\tilde{f}^{(1)}(\mathbf{x}^{(i)})}{f^{(0)}(\mathbf{x}^{(i)})}. \quad (17)$$

However, since the result of merging of (16) is in the form of empirical density this result can not be directly applied in the $j + 1$ step for two reasons:

1. the samples generated for $\tilde{f}^{(0)}$ may be in low-density regions of $\tilde{f}^{(1)}$, which calls for re-sampling.
2. values of marginal and conditional projections of $\tilde{f}^{(1)}(\mathbf{x})$ are required in the next step.

The inability to perform these steps on empirical density motivates us to approximate the resulting merger by a smooth density with known normalization.

3.2 Mixture approximation

We choose Gaussian mixtures as a suitable approximator due to their universal approximator property [9],

$$\tilde{f}(\mathbf{x}) \approx f_{\text{mix}}(\mathbf{x}) = \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Sigma_k), \quad (18)$$

where K denotes the number of components in the mixture, μ_k mean and Σ_k variance of the k th component. The simultaneous use of two distinct approximations, i.e. the empirical density and the Gaussian mixture, requires to define mutual conversions between them.

The conversion from Gaussian mixture is straightforward. If the $\tilde{f}^{(1)}$ merger is in the mixture form (18), both operations required in the previous Section are available as follows:

1. Generating samples from mixture models is a straightforward operation,
2. Marginal density $\tilde{f}(x_{s,c})$ required in (15) is again a mixture of Gaussians

$$\tilde{f}^{(j)}(x_{s,c}) = \int \tilde{f}^{(j)}(x_{s,0}, x_{s,d}, x_{s,c}) dx_{s,0} dx_{s,d} = \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_{k,s(c)}, \Sigma_{k,s(c)})$$

where $\mu_{k,s(c)}$ is a sub-vector of μ_k and $\Sigma_{k,s(c)}$ is a sub-matrix of Σ_k , each containing only elements corresponding to $x_{s,c}$.

The conditional density $\tilde{f}(x_{s,0}|x_{s,d}, x_{s,c})$ required in (15) can be computed as a ratio of marginals:

$$\tilde{f}^{(j)}(x_{s,0}|x_{s,d}, x_{s,c}) = \frac{\tilde{f}^{(j)}(\mathbf{x})}{\tilde{f}^{(j)}(x_{s,d}, x_{s,c})} = \frac{\sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Sigma_k)}{\sum_{k=1}^K \alpha_k \mathcal{N}(\mu_{k,s(d,c)}, \Sigma_{k,s(d,c)})}$$

However, conversion of (16) into a mixture is not a trivial task and it is treated in the next Section.

3.2.1 Mixture estimation

Conversion from a non-weighted empirical density to a mixture model can be approached via divergence minimization. When the minimized divergence is a Kullback-Leibler density (or Kerridge inequality), the resulting minimizer correspond to maximum likelihood [8]. This result does not hold for a weighted empirical density. We can explore two possibilities:

1. Use re-sampling to convert the weighted density to an empirical density. Use maximum likelihood estimation.
2. Derive a new minimizer for Kerridge inequality with respect to weighted empirical density.

Both approaches will be studied using the Expectation-Minimization (EM) idea [1] and their relation will be shown at the end of this Section.

Estimation of the mixture model via the EM algorithm is based on defining internal (latent) variable $l_{k,i}$ as a label if $l_{k,i} = 1$ if the i th data record was generated by the k th component of the mixture and

$l_{k,i} = 0$ otherwise. Assuming that we know that the i th data sample, x_i , was generated from the \bar{k} th component $f(x_i) = f(x_i|\mu_{\bar{k}}, \Sigma_{\bar{k}})$, the observation density of the i th sample can be rewritten using $l_{k,i}$ notation as follows:

$$f(x_i|l_i) = \prod_{k=1}^K f(x_i|\mu_k, \Sigma_k)^{l_{k,i}} \quad (19)$$

Then, the mixture model (18) arise by complementing (19) by $f(l_i) = \alpha_i$ and integrating over l_i , $l_i = [l_{1,i}, \dots, l_{K,i}]$.

The advantage of this parametrization is that estimation of parameters μ_k, Σ_k is trivial if l_i is known for each i . Then, the statistics of the posterior density $f(\mu_k, \Sigma_k|x_1, \dots, x_n)$ is of the Gauss-inverse-Wishart form with statistics V_k, ν_k which are updated as follows [11, 12]:

$$V_k = \sum_{i=1}^n l_{k,i} \begin{bmatrix} x_i \\ 1 \end{bmatrix} [x_i' \ 1] , \quad \nu_k = \sum_{i=1}^n l_{k,i}.$$

The idea of EM algorithm is to replace unknown $l_{k,i}$ by their estimates, estimate parameters μ_k, Σ_k , re-estimate $l_{k,i}$ and so on until convergence. Based on the estimate used in the algorithm we distinguish two variants:

EM-algorithm where

$$\hat{l}_{k,i} \propto f(x = x_i|\hat{\mu}_k, \hat{\Sigma}_k), \quad (20)$$

using estimates $\hat{\mu}_k, \hat{\Sigma}_k$ are parameter estimates from the previous iteration.

QB-algorithms where

$$\hat{l}_{k,i} \propto \int f(x = x_i|\mu_k, \Sigma_k) f(\mu_k, \Sigma_k) d\mu_k d\Sigma_k, \quad (21)$$

EM-like algorithm where $l_{k,\hat{i}} = 1$ for $\hat{i} = \arg \max \hat{l}_{k,i}$ where $\hat{l}_{k,i}$ is from (21). This variant correspond to an alternative formulation of EM algorithm for mixtures, where the expectation step is over the component parameters and the maximization step is with respect to the label variable. In classical EM algorithm for mixtures it is the other way around.

see [12] for more details.

In the QB case, the estimate of $\hat{l}_{k,i} \in [0, 1]$ can be interpreted as a weight of contribution of the i th data sample to the statistics of parameters μ_k and Σ_k . This observation motivates the following modification of the QB algorithm for weighted empirical densities:

1. Weights $\hat{l}_{k,i}$ are estimated using (20) or (21),
2. Statistics of the estimates of μ_k and Σ_k are updated with weight

$$l_{k,i}^* = \hat{l}_{k,i} w_i n. \quad (22)$$

I.e

$$V_k = \sum_{i=1}^n l_{k,i}^* \begin{bmatrix} x_i \\ 1 \end{bmatrix} [x_i' \ 1] , \quad S_k = \sum_{i=1}^n l_{k,i}^*.$$

Here, w_i are normalized weights (17), $\sum_{i=1}^n w_i = 1$.

Remark 4 (Soft re-sampling). Note the outlined approach offers solution to both approaches mention at the beginning of the Section, i.e. re-sampling and weighted estimation. In the case of the re-sampled empirical density for which $w_i = 1/n$, (22) reduces to the standard QB algorithm. The operation of re-sampling creates copies of samples, e.g. the j th sample is multiplied m_j times, $m_j \in [1, 2, \dots, n]$, hence the j th data record is added with weight $\sum_{i=1}^{m_j} l_{k,j} = m_j l_{k,j}$. Since m_j is a discrete quantity that should be as proportional to w_j as possible, the proposed algorithm can be seen as soft re-sampling.

3.3 Merging Algorithm

Algorithm 5. *Stochastic Merging algorithm.*

1. Choose initial proposal density $g^{(j)}(\mathbf{x})$, $j = 0$.
2. Generate samples, $\mathbf{x}^{(i)}$ from $g^{(j)}(\mathbf{x})$.
3. Estimate mixture approximation $f_{\text{mix}}^{(j)}(\mathbf{x})$ from samples $\mathbf{x}^{(i)}$.
4. Run one iteration of Algorithm 2 at points $\mathbf{x}^{(i)}$ using $\tilde{f}^{(j)}(\mathbf{x}) \equiv f_{\text{mix}}^{(j)}(\mathbf{x})$ to obtain empirical density of the merger $\tilde{f}_{\delta}^{(j+1)}(\mathbf{x})$.
5. If convergence was reached, stop.
Otherwise, compute effective sample size $\text{eff} = 1/w'w$ and
 - (a) If the effective sample size is lower than chosen threshold, goto 2.
 - (b) Otherwise, goto 3.

3.4 Open issues

Since the main merging algorithm is iterative, there are always issues related to initialization and determination of converged state. So far, we have no definite solution to these issues, only remarks on their specifics for the considered case.

3.4.1 Initialization of the proposal

The above algorithm requires the knowledge of initial proposal density g_0 , when this density is not known, it has to be designed from the sources. Note that the sources are a combination of conditional and marginal densities. If they were non-conflicting, sampling from these could be accomplished by the means of the *Gibbs sampling*. Thus, a heuristic initialization of the algorithm can use the following modification of the Gibbs sampling algorithm that can be applied even for the case of conflicting sources (however, without any guarantee of convergence):

1. Choose initial value of a sample, $x^{(0)}$
2. Set iteration counter $j = 0$,
3. Choose j th (random?) ordering of the sources, $s_1 \dots s_n$.
4. From each source s_i ,
 - (a) substitute those values of $x^{(j)}$, that are in the conditioning part of s_i
 - (b) generate new partial sample $\tilde{x}^{(i)}$ and copy its values to appropriate elements of $x^{(j)}$.
5. Increase iteration counter j , goto 3.

3.4.2 When to stop the algorithm

The original stopping rule of the the iterative algorithm proposed in Section 2.3 is based on comparing divergence of two subsequent approximations of the merged density. This variant of the stopping rule is also applicable in this case, however, the importance sampling procedure offers another measure of convergence. Since the result of merging in i th iteration is used as proposal for the next iteration, the two subsequent merging results are used to compute the importance weights (17). Thus, when the whole algorithm converges, the weights w_i should converge to 1.

The main advantage of this measure is the fact that it takes into account the influence of both approximation used in this approach—i.e. the empirical density and the mixture model—hence it may indicate even poor performance of the mixture estimation.

4 Experiments

All above mentioned algorithms were implemented in C++ toolbox BDM, which is freely available from <http://mys.utia.cas.cz:1800/trac/bdm>.

4.1 Log-normal merging

The first test of the proposed approximation is the analytical example studied in Appendix B, Fig. 4. Specifically, two one-dimensional Gaussian densities with mean values $-5, 5$ and variances $2, 10$ respectively, were tested. Two experiments with log-normal merging with $\beta = 1.2$ using analytical marginal were run with different number of samples. The first experiment with 1000 and the second with 100 samples. Each experiment was fitted with mixture of 3 Gaussian components using QB approach. The results in Fig 1 indicate that for high number of samples, the mixture model provides a very good approximation. However, for smaller number of samples, two modes of the density were not distinguished.

4.2 Merging of fragmental sources

The first experiment with fragmental sources was to test if the procedure converges to the correct answer in the case of potentially compatible sources. The sources were as follows:

$$\begin{aligned} f_1(x|y) &= \mathcal{N}(1y + 1, 0.01), \\ f_2(x) &= \mathcal{N}(2, 0.02). \end{aligned} \tag{23}$$

Note that for $y = 1$, the sources are compatible. However, y is considered to be unknown and its density will be optimized by Algorithm 2. For this simple case with predictable results, we can tested performance of the algorithm on a regular grid. Convergence of iterations in Algorithm 2 for 20×20 regular grid $x \in \langle 0, 4 \rangle, y \in \langle -1, 3 \rangle$ is illustrated in Figure 2 via mean and variance of the joint merger $f(\mathbf{x})$.

Note that evaluation on the grid, even with lower number of evaluation points (400 for grid and 1000 for importance sampling) converges faster to the expected result, $\hat{y} = 1, \hat{x} = 2$. Moreover, the importance sampling procedure does not converge exactly to the expected value, but to biased values $\hat{y} = 1.05, \hat{x} = 2.03$. The exact value of the bias depends on random seed used in the procedure.

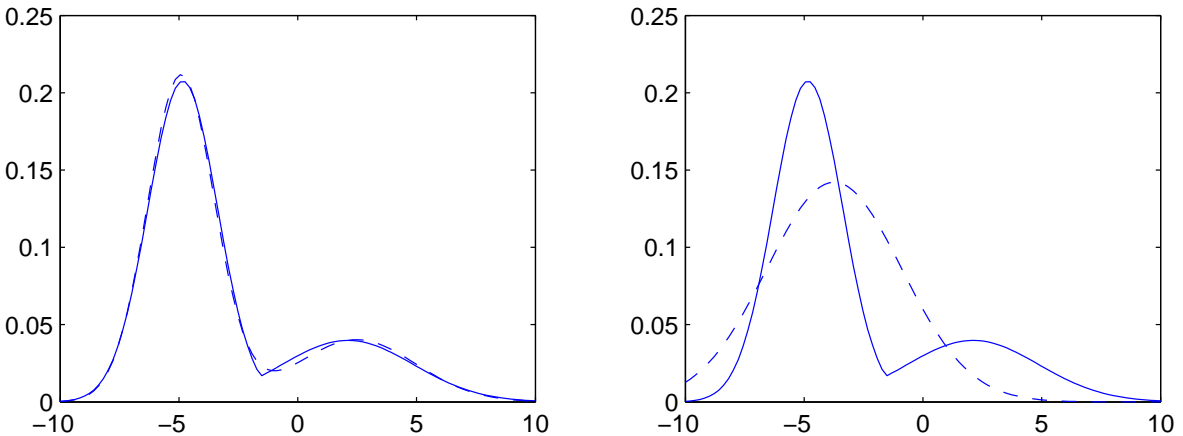


Figure 1: Approximate log-normal merging of two Gaussian densities. Solid line denote analytical merging formula evaluated on a grid. Dashed line denote the fitted mixture model evaluated on the same grid. **Left**: experiment for 1000 samples. **Right**: experiment for 100 samples.

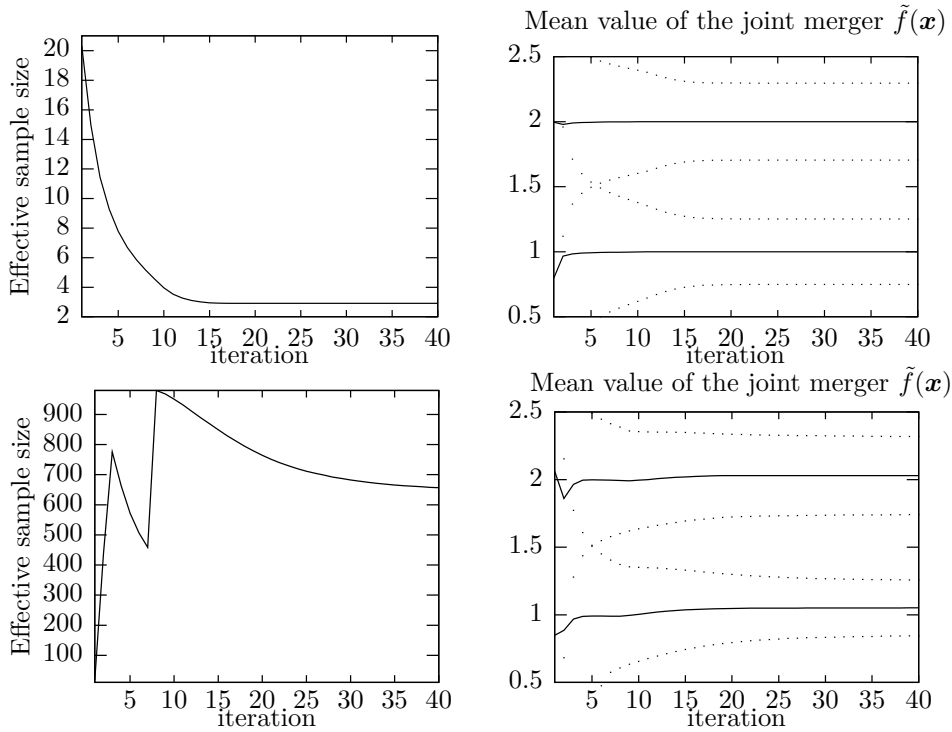


Figure 2: Convergence of the joint merger for sources (23) on discrete 20×20 grid (top row) and using importance sampling procedure (bottom row). **Left:** effective sample size, ($\text{eff} = N_s/w'w$), **Right:** mean value with 2std bounds of the joint merger $\tilde{f}(\mathbf{x})$.

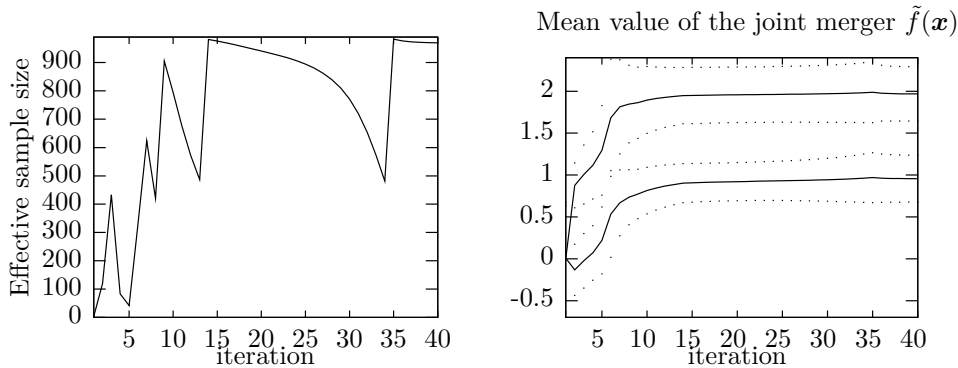


Figure 3: Convergence of the joint merger for sources (23) using importance sampling procedure when initiated far from the optimum.

Offset can be obtained even for evaluation on the grid, e.g. when the grid is not symmetrical around the expected results, or if it does not cover some parts of the space where the expected merger has significant values. Note also, that the direction of convergence of the mean value changes after re-sampling (sharp change on the effective sample size graph). This suggests to modify Algorithm 5 to allow Algorithm 2 reach steady-state.

Remark 6. Modification of Algorithm 5 which performs re-sampling only when Algorithm 2 reaches steady-state. Note from Fig. 23, that effective sample size is dropping during convergence. Before reaching steady-state, the effective sample size may reach very low values inducing numerical instability of the procedure. In cases when the initial guess of the merger was far away from the converged value, such numerical instability occurred regularly.

Significant advantage of the importance sampling procedure is the ability to converge to the correct location even when from distant initial density. Convergence of the procedure with initialization at $\mathcal{N}([0, 0], 0.1I_2)$ is displayed in Figure 3.

4.3 Merging of conditional fragmental sources

The final merger from the previous experiment converged to the expected position in which the sources were almost consistent. In this experiment, we attempt to merge sources which are consistent for more than one point. Sources

$$\begin{aligned} f_1(x|y) &= \mathcal{N}(y + 1, 0.01), \\ f_2(x|z) &= \mathcal{N}(2z - 1, 0.01), \end{aligned}$$

are consistent for any $y = 2z - 2$. Since the sources do not contain any information about the mutual relation of y and z , we expect that the iterative algorithm will preserve the relation of the initial density. For example, for choosing $y = z$ the expected value of the merger is $[3, 2, 2]$, while for $y = -z$ the expected merger has mean value at $[\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}]$.

A Monte Carlo study of 100 runs of the importance sampling algorithm with $\beta = 1, c = 1, n = 2000$ and 40 iterations was run for two initial densities:

$$\tilde{f}_A^{(0)}(x, y, z) = \mathcal{N} \left(\begin{bmatrix} 1.5 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & & \\ & 1 & 0.95 \\ & 0.95 & 1 \end{bmatrix} \right), \tilde{f}_B^{(0)}(x, y, z) = \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & \\ & 1 & -0.95 \\ & -0.95 & 1 \end{bmatrix} \right),$$

the former favoring $y \approx z$, the latter $y \approx -z$. Results of the Monte Carlo for mean value of the joint merger are in the following table:

	Monte Carlo mean	MC median	MC std
estimates for $f_A^{(0)}$	[2.21,1.20,1.66]	[2.59,1.59,1.79]	[1.39,1.40,0.52]
estimates for $f_B^{(0)}$	[0.30,-0.70,0.65]	[0.33,-0.66,0.66]	[0.25,0.26,0.01]

In most of the cases, the converged mean of the merger is where expected, which is demonstrated by the median. However, switches or slow shifts towards another points were frequent which shifted the mean value and caused large standard deviation. In this experiment, the number of iterations was set to 40 since convergence of the algorithm was hard to assess. The merger was slowly wandering in the whole space.

Results of the same example were tested for the grid-based approach, Remark 3. The results were heavily dependent on the chosen grid. If the grid was too coarse, rapid switches between various points were common. A finer grid of e.g. 40 points in one direction achieved smoother results, however, it significantly increased the number of necessary computations.

5 Conclusion

The log-normal merging of probability densities proposed by [5] was studied. The main results are:

1. Analytical solution for posterior expectation of the log-normal ‘hyper’-model.
2. Evaluation procedure of the merging using importance sampling and mixture of Gaussians.

The first result is achieved when the number of sources S is known. We have shown that properties of the log-normal merging heavily depends on the chosen prior quantity $\beta \in [1, \infty]$. Boundaries of the admissible range of β were found to correspond to two classical approaches: arithmetic (only approximately) and geometric merging, respectively. It can be argued that these boundaries are extremes and β should be chosen slightly off the boundaries.

The new method of evaluation using importance sampling was tested in simulation and the results compared with numerical evaluation on a rectangular grid. The conclusions are:

- Evaluation of the merging procedure on the grid appears to be a better option for smaller dimensional densities with well known and limited area of support with high density. The main advantage of this approach is simplicity of implementation, fast convergence, no additional tuning knobs.
- Evaluation using importance sampling with mixture models appears to be more suitable for higher dimensional densities and for sources where prediction of high-density regions is hard or impractical. The approach is still in early stage of development, with many tuning knobs to be chosen especially in the mixture estimation procedure. Also, assessment of convergence of the iterative algorithm is more demanding than in the case of grid evaluation.
- The iterative algorithm seems to converge to the expected solution in ‘simple’ cases with well predictable solution. However, convergence may be problematic in more demanding cases with highly uninformative sources. Therefore, the algorithm should be used carefully in such situations.

The algorithm can be still significantly improved. Specifically in on-line scenarios where repetitive merging of similar densities will be performed, the samples and mixture parameters from the previous step can be used to initialize the procedure in the subsequent step.

A Log-Normal Merging

Analytical marginal of the log-normal model is computed here using symbolic software maple. Notation of the output is slightly changed; here, the number of sources is denoted ν instead of S which was used in the main text.

```
> restart:with( RealDomain ):assume(g::real,g>0);
```

```
\QTR{cstyle10}{Warning, these protected names have been redefined and unprotected:
Im, Re, '\symbol{94}', arccos, arccosh, arccot, arccoth, arccsc,
arccsch, arcsec, arcsech, arcsin, arcsinh, arctan, arctanh, cos, cosh,
cot, coth, csc, csch, eval, exp, expand, limit, ln, log, sec, sech,
signum, simplify, sin, sinh, solve, sqrt, surd, tan, tanh}\QTR{cstyle10}{}
```

```
Observation model - lognormal density on g
```

```
> assume(mu::real);assume(sigma::real,sigma>0);
```

```
> f_nn:=1/d*exp(-1/2*(log(d)-mu)^2/sigma);
```

$$f_{nn} := \frac{e^{(-1/2 \frac{(\ln(d)-\mu)^2}{\sigma})}}{d}$$

```
> nk:=int(f_nn,d=0..infinity);
```

$$nk := \sqrt{\pi} \sqrt{2} \sqrt{\sigma}$$

```
> f:=f_nn/nk;
```

$$f := \frac{1}{2} \frac{e^{(-1/2 \frac{(\ln(d) - \mu^-)^2}{\sigma^-})} \sqrt{2}}{d \sqrt{\pi} \sqrt{\sigma^-}}$$

Its expected value

> Eg:=int(d*f,d=0..infinity);

$$Eg := e^{(\mu^- + 1/2 \sigma^-)}$$

Posterior on parameters - Normal-Gamma density

> assume(nu::real,nu>2);assume(lambda::real,lambda>0);assume(sl::real);
 > assume(beta::real,beta>0);
 > fms_nn:=sigma^(-1/2*nu)*exp(-1/2*1/sigma*(nu*(sl-mu)^2+lambda))*exp(-
 > beta*(sigma));

$$fms_nn := \sigma^{-(-1/2 \nu^-)} e^{(-1/2 \frac{\nu^- (sl^- - \mu^-)^2 + \lambda^-}{\sigma^-})} e^{(-\beta^- \sigma^-)}$$

> simplify(Eg*fms_nn);

$$e^{(-1/2 \frac{-2 \mu^- \sigma^- - \sigma^{-2} + \nu^- sl^{-2} - 2 \nu^- sl^- \mu^- + \nu^- \mu^{-2} + \lambda^- + 2 \beta^- \sigma^{-2}}{\sigma^-})} \sigma^{-(-1/2 \nu^-)}$$

> nk_ms:=int(int(fms_nn,mu=-infinity..infinity),sigma=0..infinity):
 > fms:=fms_nn/nk_ms:

Its marginals

> fm:=simplify(int(fms,sigma=0..infinity)):

> fs:=simplify(int(fms,mu=-infinity..infinity)):

Expectations of marginals

> Ms:=solve(diff(fs,sigma)=0,sigma);

$$Ms := \frac{1}{4} \frac{-\nu^- + 1 + \sqrt{\nu^{-2} - 2\nu^- + 1 + 8\beta^- \lambda^-}}{\beta^-}, -\frac{1}{4} \frac{\nu^- - 1 + \sqrt{\nu^{-2} - 2\nu^- + 1 + 8\beta^- \lambda^-}}{\beta^-}$$

> with(student):completesquare(Ms[01],nu);

$$\frac{1}{4} \frac{-\nu^- + 1 + \sqrt{(\nu^- - 1)^2 + 8\beta^- \lambda^-}}{\beta^-}$$

> fmc:=fms/fs:

> fmc:=simplify(fmc);

$$fmc := \frac{1}{2} \frac{\sqrt{2} e^{(-1/2 \frac{\nu^- (sl^- - \mu^-)^2}{\sigma^-})} \sqrt{\nu^-}}{\sqrt{\sigma^-} \sqrt{\pi}}$$

> Emu:=(int(mu*fmc,mu=-infinity..infinity));

$$Emu := sl^-$$

> Ed:=int(Eg*fmc,mu=-infinity..infinity);

$$Ed := e^{(1/2 \frac{\sigma^- \nu^- + \sigma^- + 2 \nu^- sl^-}{\nu^-})}$$

> Eds:=simplify(subs(sigma=Ms[1],Ed));

$$Eds := e^{(1/8 \frac{-\nu^{-2} + \nu^- \sqrt{\nu^{-2} - 2\nu^- + 1 + 8\beta^- \lambda^-} + 1 + \sqrt{\nu^{-2} - 2\nu^- + 1 + 8\beta^- \lambda^-} + 8\nu^- sl^- \beta^-}}{\beta^- \nu^-}}$$

> Eda:=int(Ed*fs,sigma=0..infinity);

$$\begin{aligned}
Eda := & 2^{(1/4\nu^- - 5/4)} e^{st^-} \lambda^{-(5/4 - 1/4\nu^-)} \beta^{-(5/4 - 1/4\nu^-)} \left(\right. \\
& \frac{2^{(5/2 - 1/2\nu^-)} (\nu^- - 1 + \beta^- (2 - \frac{\nu^- + 1}{\nu^- \beta^-}) \lambda^-) \%1^{(1/2\nu^- + 1/2)} \text{BesselK}(\frac{1}{2}\nu^- + \frac{1}{2}, \%1)}{\beta^{-2} (2 - \frac{\nu^- + 1}{\nu^- \beta^-})^2 \lambda^{-2}} \\
& \left. - \frac{2^{(5/2 - 1/2\nu^-)} (\nu^- - 1) \%1^{(1/2\nu^- + 1/2)} \text{BesselK}(\frac{1}{2}\nu^- + \frac{3}{2}, \%1)}{\beta^{-(3/2)} (2 - \frac{\nu^- + 1}{\nu^- \beta^-})^{(3/2)} \lambda^{-(3/2)}} \right) / (\\
& \sqrt{\lambda^-} \sqrt{\beta^-} \text{BesselK}(\frac{1}{2}\nu^- + \frac{1}{2}, \sqrt{\beta^-} \sqrt{2} \sqrt{\lambda^-}) \sqrt{2} - \nu^- \text{BesselK}(\frac{1}{2}\nu^- - \frac{1}{2}, \sqrt{\beta^-} \sqrt{2} \sqrt{\lambda^-}) \\
& + \text{BesselK}(\frac{1}{2}\nu^- - \frac{1}{2}, \sqrt{\beta^-} \sqrt{2} \sqrt{\lambda^-}) \\
& \%1 := \sqrt{\beta^-} \sqrt{2 - \frac{\nu^- + 1}{\nu^- \beta^-}} \sqrt{\lambda^-}
\end{aligned}$$

Greatly simplified when nu is known

> Eda_nu2:=simplify(subs({nu=2},Eda));

$$Eda_nu2 := 2 \frac{\sqrt{\frac{4\beta^- - 3}{\beta^-}} e^{(-1/2 \sqrt{\beta^-} \sqrt{2} \sqrt{\frac{4\beta^- - 3}{\beta^-}} \sqrt{\lambda^-} + st^- + \sqrt{\beta^-} \sqrt{2} \sqrt{\lambda^-})} \sqrt{\sqrt{\beta^-} \sqrt{\lambda^-}} \beta^{-(3/4)}}{\lambda^{-(1/4)} (4\beta^- - 3)}$$

> Eda_n2_b1:=simplify(subs({beta=1},Eda_nu2));

$$Eda_n2_b1 := 2 e^{(st^- + 1/2 \sqrt{2} \sqrt{\lambda^-})}$$

> Eda_n2_b2:=simplify(subs({beta=2},Eda_nu2));

$$Eda_n2_b2 := \frac{2}{5} \sqrt{5} \sqrt{2} e^{(st^- - 1/2 \sqrt{2} \sqrt{5} \sqrt{\lambda^-} + 2 \sqrt{\lambda^-})}$$

> Eds_n2:=subs(nu=2,Eds);

$$Eds_n2 := e^{(1/16 \frac{-3+3\sqrt{1+8\beta^- \lambda^-} + 16 st^- \beta^-}{\beta^-})}$$

B On the choice of prior

The result of log-normal merging heavily depend on the choice of coefficient β in prior density (9). This coefficient is necessary to compensate the term $\exp(\frac{1}{2}\sigma)$ in (10) which is increasing with $\sigma \rightarrow \infty$ yielding infinite value of the merged density. This term is compensated by $\beta = \frac{1}{2}$, however it is still not sufficiently compensated for the case with two sources,

$$E(d) = 2 \frac{\sqrt{\frac{4\beta - 3}{\beta}} \beta \exp\left(-\frac{1}{2} \sqrt{\beta} \sqrt{2} \sqrt{\frac{4\beta - 3}{\beta}} \sqrt{\lambda} + \hat{\mu} + \sqrt{\beta} \sqrt{2} \sqrt{\lambda}\right)}{(4\beta - 3)},$$

hence, $\beta > \frac{3}{4}$ in order to obtain acceptable solution. Similarly, for four sources $\beta > \frac{5}{8}$, from which we can extrapolate that $\beta > \frac{S+1}{2S}$. This condition was however derived only to assure finite values of $\tilde{f}(\hat{x})$ at point \hat{x} , without any attempt to normalize the resulting $\tilde{f}(x)$ on its support. This may be difficult since

$$\lim_{\beta \rightarrow \frac{S+1}{2S}} E(d) \rightarrow \infty.$$

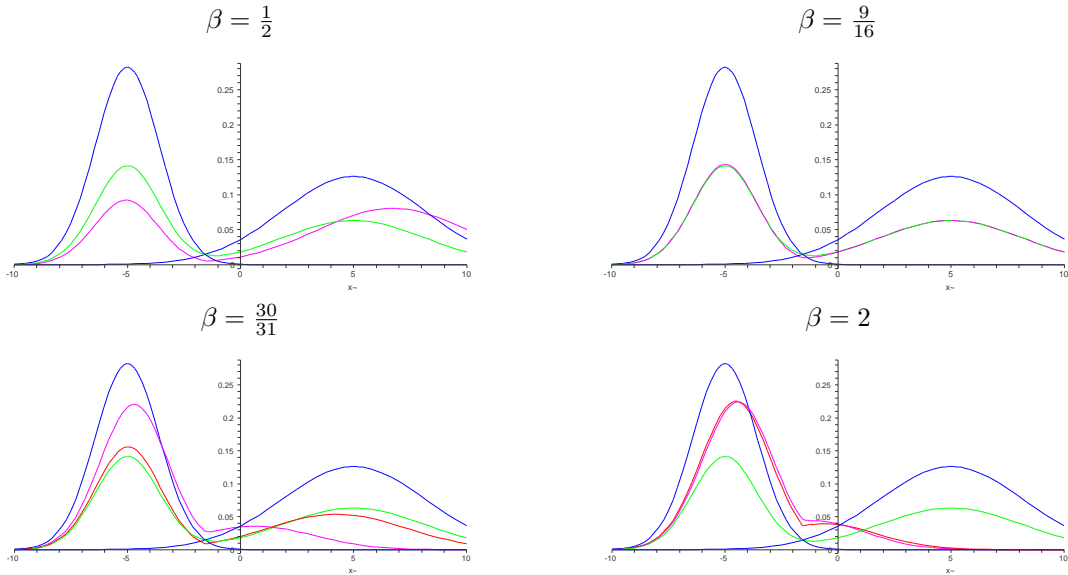


Figure 4: Merging of two one-dimensional Gaussian sources for different values of β , from top to bottom $\beta = [\frac{2}{3}, \frac{31}{30}, 2]$. Blue lines denote source pdfs, green line is the arithmetic merger, red is the marginalized (10) and magenta the conditioned (13) merger.

In order to address normalization, we would have to substitute source densities into the final formula for $E(d)$. A range of experiment was performed with two Gaussian sources. The results are summarized in Figure 4. We note the following:

- For β below the analytical limit, the exact marginal is not well defined, however, the conditional estimate (13) is still well defined. The result of conditional merger for $\beta = \frac{1}{2}$ is undesired, since it favors low density areas.
- The conditional merger for $\beta = \frac{9}{16}$ is almost exactly the arithmetic merging, for lower β it has the effect as with $\beta = \frac{1}{2}$, for higher β it is narrowing the resulting density.
- Numerical normalization of the resulting merger was possible only for $\beta > 1$ and was getting increasingly difficult around 1, that is why the plot in figure 4 is at value $\frac{31}{30}$.

Hence, the log-normal approach to merging offers a smooth transition between the arithmetic and geometric merging. In the exact case, the arithmetic merger arise for $\beta = 1$ (at least for $S = 2$) and the geometric merger arise for $\beta \rightarrow \infty$.

Acknowledgment:

Support of grants MŠMT 1M0572 and GAČR 102/08/P250 is gratefully acknowledged.

References

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B*, 39:1–38, 1977.
- [2] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

- [3] Ch. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1:114–148, 1986.
- [4] M. Kárný and T.V. Guy. On dynamic decision-making scenarios with multiple participants. In J. Andřýsek, M. Kárný, and J. Kracík, editors, *Multiple Participant Decision Making*, pages 17–28, Adelaide, May 2004. Advanced Knowledge International.
- [5] M. Kárný, T.V. Guy, A. Bordini, and F. Rugeri. Cooperation via sharing of probabilistic elements. Technical Report 9-MI, IMATI Milano, 2008.
- [6] J. Kracík. Composition of Probability Density Functions - Optimizing Approach. Technical Report 2099, ÚTIA AV ČR, Praha, 2004.
- [7] J. Kracík. On composition of probability density functions. In J. Andřýsek, M. Kárný, and J. Kracík, editors, *Multiple Participant Decision Making*, pages 113–121, Adelaide, May 2004. Advanced Knowledge International.
- [8] R. Kulhavý. A geometric approach to statistical estimation. In *Proceedings of the 34th IEEE Conference on Decision and Control*, volume 2, pages 1097–1102, New Orleans, LA, 1995.
- [9] V. Maz`ya and G. Schmidt. On approximate approximations using Gaussian kernels. *IMA Journal of Numerical Analysis*, 16(1):13–29, 2001.
- [10] P. A. Morris. Combining expert judgments: A bayesian approach. *Management Science*, 23:679–693, 1977.
- [11] V. Peterka. Bayesian approach to system identification. In P. Eykhoff, editor, *Trends and Progress in System identification*, pages 239–304. Pergamon Press, Oxford, 1981.
- [12] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2005.
- [13] M. West and J. Crosse. Modelling probabilistic agent opinion. *Journal of the Royal Statistical Society. Series B*, 54:285–299, 1992.