

Bayesian Inference for the Weights in Logarithmic Pooling

Luiz M. Carvalho*, Daniel A. M. Villela†, Flavio C. Coelho*, and Leonardo S. Bastos†

Abstract. Combining distributions is an important issue in decision theory and Bayesian inference. Logarithmic pooling is a popular method to aggregate expert opinions by using a set of weights that reflect the reliability of each information source. However, the resulting pooled distribution depends heavily on set of weights given to each opinion/prior and thus careful consideration must be given to the choice of weights. In this paper we review and extend the statistical theory of logarithmic pooling, focusing on the assignment of the weights using a hierarchical prior distribution. We explore several statistical applications, such as the estimation of survival probabilities, meta-analysis and Bayesian melding of deterministic models of population growth and epidemics. We show that it is possible learn the weights from data, although identifiability issues may arise for some configurations of priors and data. Furthermore, we show how the hierarchical approach leads to posterior distributions that are able to accommodate prior-data conflict in complex models.

Keywords: logarithmic pooling, expert opinion, hierarchical modelling, Bayesian melding.

MSC2020 subject classifications: Primary 60K35, 60K35; secondary 60K35.

1 Introduction

Combining probability distributions is a topic of general interest, both in the statistical (West, 1984; Genest et al., 1986; Genest and Zidek, 1986) and decision theory literatures (Genest et al., 1984; French, 1985; Pennock and Wellman, 1997; Guardoni, 2002). On the theoretical front, studying opinion pooling operators may give important insights on consensus belief formation and group decision making (West, 1984; Genest and Zidek, 1986; Guardoni, 2002). Among the various opinion pooling operators proposed in the literature, logarithmic pooling has enjoyed much popularity, mainly due to its many desirable properties such as relative propensity consistency (RPC) and external Bayesianity (EB) (Genest et al., 1986) – see Remark 2.1 below. In a practical setting, logarithmic pooling finds use in a wide range of fields, from engineering (Lind and Nowak, 1988; Savchuk and Martz, 1994) to wildlife conservation (Poole and Raftery, 2000) and infectious disease modelling (Coelho and Codeço, 2009).

A common situation of interest is combining expert opinions about a quantity of interest $\theta \in \Theta \subseteq \mathbb{R}^p$ when these opinions can be represented as (proper) probability

arXiv: 1502.04206

*School of Applied Mathematics, Getúlio Vargas Foundation (FGV), Rio de Janeiro, Brazil, lmax.fgv@gmail.com, daniel.villela@fiocruz.br

†Scientific Computing Programme, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil, fccoelho@gmail.com, lsbastos.fiocruz@gmail.com

distributions. Combining these opinions using logarithmic pooling requires assigning weights to each of the experts, which represent the (relative) reliability of each opinion (Genest et al., 1984; French, 1985). This requirement naturally leads to the question of how to choose the weights in a meaningful way, according to some well-accepted (optimality) criterion. There are a few proposals in the literature that build methods using different approaches. One proposal is to maximise the entropy of the pooled distribution (Myung et al., 1996), whereas another one is to minimise Kullback-Leibler (KL) divergence between the pooled distribution and the individual opinions (Abbas, 2009) or between the pooled (prior) distribution and the posterior distribution (Rufo et al., 2012a,b).

While moving away from the problem of arbitrarily assigning the weights, these approaches arrive at single point solutions, similar to point estimates in statistical theory. While we acknowledge that these approaches have merit, we argue that in many settings it would be desirable to incorporate information on the relative reliabilities of the experts into the pooling procedure while accommodating uncertainty about the weights. Moreover, assigning a probability distribution over the weights allows one to obtain a posterior distribution using a Bayesian procedure, which in turn enables learning about the weights from data (Poole and Raftery, 2000). Therefore, it makes possible to sequentially update knowledge about the reliability of each expert/source in the face of new data.

In this paper we discuss previous approaches for assigning the weights based on optimality criteria and study the issue of assigning hierarchical priors to the weights in order to learn about them from data. This paper is organised as follows: in Section 2 we introduce the necessary concepts and notation on logarithmic pooling, as well as some of its key properties. We also prove a new result about log-concavity of the pooled distribution when all distributions are log-concave. In Section 3 we present different approaches to choosing the weights, two methods based on optimality criteria, namely maximising the entropy of the pooled prior and minimising Kullback-Leibler divergence between the pooled distribution and the expert distributions. In addition we also lay out an approach hierarchical modelling of the weights. Section 4 contains applications of logarithmic pooling to reliability analysis (Sections 4.1 and 4.2) and Bayesian melding (Section 4.3). We conclude with a discussion of our results in light of the statistical literature in Section 5.

2 Logarithmic pooling: properties and applications

In this section we introduce the necessary theory and notation and motivate the use of the logarithmic pooling operator by presenting some of its desirable properties.

First let us define the logarithmic pooling (LP) operator. Let $\mathbf{F}_\theta := \{f_0(\theta), f_1(\theta), \dots, f_K(\theta)\}$ be a set of (densities of) distributions representing the opinions of $K + 1$ experts and let $\boldsymbol{\alpha} := \{\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_K\}$ be the vector of weights, such that $\alpha_i \geq 0 \forall i$ and

$\sum_{i=0}^K \alpha_i = 1$. The log-pooled density is

$$\mathcal{LP}(\mathbf{F}_\theta, \boldsymbol{\alpha}) := \pi(\theta \mid \boldsymbol{\alpha}) = t(\boldsymbol{\alpha}) \prod_{i=0}^K f_i(\theta)^{\alpha_i}, \tag{2.1}$$

where $t(\boldsymbol{\alpha}) = \left[\int_{\Theta} \prod_{i=0}^K f_i(\theta)^{\alpha_i} d\theta \right]^{-1}$.

Logarithmic pooling will only yield proper probability distributions if it is possible to normalise the expression in (2.1). This condition is usually assumed implicitly, without proof. While Poole and Raftery (2000) provide a proof for the case of two densities (see Theorem 1 therein), Genest et al. (1986, p. 489) prove the result for a finite number of densities:

Theorem 2.1 (Normalisation (Genest et al., 1986)). *Let \mathcal{A} be a $(K + 1)$ -dimensional open simplex on $[0, 1]$. For all $\boldsymbol{\alpha} \in \mathcal{A}$ there exists a constant $t(\boldsymbol{\alpha})$ such that $\int_{\Theta} \pi(\theta \mid \boldsymbol{\alpha}) d\theta = 1$.*

We give a simple proof using Hölder’s inequality in Appendix A (Carvalho et al., 2022). This result ensures any (finite) number of proper distributions can be combined using the logarithmic pooling operator to yield a normalisable (proper) density. In addition, log-linear pools enjoy the *external Bayesianity* property (Remark 2.1), which guarantees that whether one combines the expert opinions before or after observing evidence does not affect the resulting pooled distribution.

Remark 2.1 (External Bayesianity (Genest, 1984; Genest et al., 1984)). If the expert opinions are given by densities $f_i(\theta)$ and one observes data x such that one can specify a likelihood $l(x \mid \theta)$, combining the set of posteriors $p_i(\theta \mid x) \propto l(x \mid \theta)f_i(\theta)$ yields the same distribution as combining the densities f_i to obtain a prior $\pi(\theta)$ and then combine it with $l(x \mid \theta)$ to obtain a posterior $p(\theta \mid x) \propto l(x \mid \theta)\pi(\theta)$.

The proof of this fact is omitted here. Genest et al. (1984) show that the logarithmic pooling operator in (2.1) is the **only** aggregation (pooling) operator that enjoys external Bayesianity. Moreover, the logarithmic pooling operator has the relative propensity consistency (RPC) property (Remark 2.2), whereby the pooled opinion preserves relative judgments from the experts.

Remark 2.2 (Relative propensity consistency (Genest et al., 1984)). Taking \mathbf{F}_X as a set of expert opinions with support on a space \mathcal{X} , define $\boldsymbol{\xi} = \{\mathbf{F}_X, a, b\}$ for arbitrary $a, b \in \mathcal{X}$. Let \mathcal{T} be a pooling operator and define two functions U and V such that

$$U(\boldsymbol{\xi}) := \left(\frac{f_0(a)}{f_0(b)}, \frac{f_1(a)}{f_1(b)}, \dots, \frac{f_K(a)}{f_K(b)} \right) \text{ and } V(\boldsymbol{\xi}) := \frac{\mathcal{T}_{\mathbf{F}_X}(a)}{\mathcal{T}_{\mathbf{F}_X}(b)}.$$

We then say that \mathcal{T} enjoys *relative propensity consistency* (RPC) if and only if

$$U(\boldsymbol{\xi}_1) \geq U(\boldsymbol{\xi}_2) \implies V(\boldsymbol{\xi}_1) \geq V(\boldsymbol{\xi}_2),$$

for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$, with inequalities taken element-wise where necessary.

We refer the reader to Genest et al. (1984) for a proof. Informally, this property says that if all experts consider a particular event A more probable than another event B , then the pooled opinion should be consistent with these relative judgments. Genest et al. (1984) show that for mild conditions on \mathcal{X} , namely $|\mathcal{X}| \geq 3$, the logarithmic pooling operator is the only pooling operator with RPC (see also Lemma A.1 in Appendix A (Carvalho et al., 2022)).

Another desirable property of the logarithmic pooling operator is log-concavity. Log-concavity of the pooled prior may be important to consider in order to guarantee unimodality and certain conditions on tail behaviour – see Bagnoli and Bergstrom (2005). Certain algorithms, such as slice sampling (Neal, 2003), also rely on log-concavity of the target distribution. See Saumard and Wellner (2014) for a review of the applicability of log-concavity in Statistics.

These considerations motivate the following theorem, which is, to the best of our knowledge, a new result:

Theorem 2.2 (Log-concavity). *Under the mild regularity conditions mentioned in Remark 2.2, if \mathbf{F}_θ is a set of log-concave distributions, then $\pi(\theta | \boldsymbol{\alpha})$ is also log-concave. Moreover, logarithmic pooling is the only pooling operator that will always produce a log-concave density when all the elements of \mathbf{F}_θ are log-concave.*

Proof. See Appendix A (Carvalho et al., 2022). □

Theorem 2.2 tells us that logarithmic pooling is the only aggregation method to universally preserve log-concavity, for any configuration of the weights ($\boldsymbol{\alpha}$). This universality result is important because it holds for any set of log-concave distributions, \mathbf{F}_θ . In contrast, a linear pool of $K + 1$ Gaussian distributions with common mean, $\pi_{\text{linear}}(\theta) = \sum_{i=0}^K \alpha_i f_i(\theta)$, would produce a log-concave pooled distribution for any $\boldsymbol{\alpha}$, but this would potentially fail if the means were different.

2.1 Exponential family

The exponential family of probability distributions finds widespread use in the modelling of empirical phenomena (DasGupta, 2011). In this section we give expressions for the entropy and Kullback-Leibler divergence for the pooled distributions. These will be useful in applications presented later in the paper.

Suppose we are interested in a random quantity θ whose distribution belongs to the exponential family with parameter ψ and probability density function (pdf) given by (Robert, 2007, p. 115):

$$f(\theta|\psi) = h(\theta)e^{\psi\theta - s(\psi)}. \quad (2.2)$$

Let \mathbf{F}_θ be a set of densities on θ of the form in (2.2), $f_i(\theta|\psi_i)$, $i = 0, 1, \dots, K$. The combined (log-pooled) distribution also belongs to the exponential family:

$$\pi(\theta|\boldsymbol{\alpha}) = t(\boldsymbol{\alpha})h^*(\theta)e^{\psi^*y - s^*(\psi)}, \quad (2.3)$$

where $\boldsymbol{\psi} := \{\psi_0, \psi_1, \dots, \psi_K\}$, $h^*(\boldsymbol{\theta}) = \prod_{i=0}^K h_i(\boldsymbol{\theta})^{\alpha_i}$, $\boldsymbol{\psi}^* = \sum_{i=0}^K \alpha_i \psi_i$ and $s^*(\boldsymbol{\psi}) = \sum_{i=0}^K \alpha_i s_i(\psi_i)$.

The entropy function of the log-pooled distribution is

$$H_\pi(\boldsymbol{\theta}; \boldsymbol{\alpha}) := -\mathbb{E}_\pi[-\log \pi(\boldsymbol{\theta}|\boldsymbol{\alpha})] = -\log t(\boldsymbol{\alpha}) + s^*(\boldsymbol{\psi}) - \mathbb{E}_\pi[\log h^*(\boldsymbol{\theta})] - \boldsymbol{\psi}^* \mathbb{E}_\pi[\boldsymbol{\theta}], \quad (2.4)$$

where $\mathbb{E}_\pi[g(\boldsymbol{\theta})]$ is the expectation of a π -measurable function $g(\cdot)$ with respect to $\pi(\boldsymbol{\theta}|\boldsymbol{\alpha})$, when the integral exists.

The Kullback-Leibler divergence between the pooled distribution (2.3) and each distribution in \mathbf{F}_y can be written as:

$$KL(\pi||f_i) = -H_\pi(\boldsymbol{\theta}; \boldsymbol{\alpha}) - \mathbb{E}_\pi[\log h_i(\boldsymbol{\theta})] - \boldsymbol{\theta}_i \mathbb{E}_\pi[\boldsymbol{\theta}] + s_i(\psi_i).$$

These expressions allow for easy computation of information measures for a broad class of distributions, which will be useful in the remainder of this paper (see also Appendix E, Carvalho et al., 2022).

Conjugate priors to the exponential family

Here we can see $f(\boldsymbol{\theta} | \boldsymbol{\psi})$ as a ‘‘likelihood’’ and this motivates thinking about conjugate prior measures for $\boldsymbol{\psi}$ as a means of specifying hierarchical priors on all quantities of interest. A conjugate prior family for $f(\boldsymbol{\theta} | \boldsymbol{\psi})$ (2.2), has the following form (Diaconis and Ylvisaker, 1979):

$$g(\boldsymbol{\psi}|a, b) = C(a, b)e^{\boldsymbol{\psi}a - b s(\boldsymbol{\psi})}, \quad (2.5)$$

where $C(a, b)$ is a normalising constant. Similar to the above, let $\mathbf{G}_\boldsymbol{\psi}$ be a set of log-conjugate prior distributions representing the opinions of $K + 1$ experts, and $g_i(\boldsymbol{\psi}) = g(\boldsymbol{\psi}|a_i, b_i)$ from equation (2.5).

The log-pooled prior is also a conjugate prior for $f(\boldsymbol{\theta} | \boldsymbol{\psi})$ with hyperparameters given by a weighted mean of the experts’s hyperparameters, i.e., $\pi(\boldsymbol{\psi} | \boldsymbol{\alpha}) = g(\boldsymbol{\psi} | a^*, b^*)$, where $a^* = \sum_{i=0}^K \alpha_i a_i$ and $b^* = \sum_{i=0}^K \alpha_i b_i$.

The entropy function of the log-pooled prior (2.5) is given by

$$H_\pi(\boldsymbol{\psi}; \boldsymbol{\alpha}) = -\log(C(a^*, b^*)) - a^* \mathbb{E}_\pi[\boldsymbol{\psi}] + b^* \mathbb{E}_\pi[s(\boldsymbol{\psi})]. \quad (2.6)$$

And the Kullback-Leibler divergence, $KL(\pi||g_i)$, is the following

$$KL(\pi||g_i) = -H_\pi(\boldsymbol{\psi}; \boldsymbol{\alpha}) - \log(C(a_i, b_i)) - a_i \mathbb{E}_\pi[\boldsymbol{\psi}] + b_i \mathbb{E}_\pi[s(\boldsymbol{\psi})]. \quad (2.7)$$

2.2 Known limitations

Despite its many attractive properties, the logarithmic pool also presents limitations that might hinder its application in practice. In this section we briefly review these limitations in order to prepare the reader for how they manifest later in the paper.

We begin with a known quirk of logarithmic pools: The support of the pooled distribution is $\text{supp}(\pi) = \bigcap_{i=0}^K \text{supp}(f_i)$, i.e., the pooled distribution will have the smallest support amongst the distributions being combined. This means a single expert can make large portions of the sample space impossible under the pooled distribution. In the extreme case where one of the opinions is a point-mass, the pool will also be a point-mass, not matter the weights and the Again, however, the analyst can use external considerations to exclude an expert whose probability density has too narrow a support. As we will see later on, this property manifests subtly in related situation which is when one of the experts elicits a very concentrated probability distribution, encoding a very precise opinion.

Logarithmic pools are also hard to compute exactly for arbitrary collections of opinions \mathbf{F}_θ because the exponential dependence on the weights leads to difficulties computing the normalising constant $t(\boldsymbol{\alpha})$. In situations where $\boldsymbol{\alpha}$ is random (see Section 3.2) one cannot bypass computing the normalising constant (Neuenschwander et al., 2009) and this adds a layer of computational complexity – see Discussion.

3 Assigning the weights in logarithmic pooling

The weights ($\boldsymbol{\alpha}$) play a key role on the logarithmic pooling and hence their choice is critical. Building on work by Poole and Raftery (2000); Rufo et al. (2012a,b) and Abbas (2009), we now move on to study three approaches to assigning the weights in logarithmic pooling. The first two approaches are based on optimality criteria and a third method proposes assigning a (hyper)prior to the weights.

3.1 Choosing weights based on optimality criteria

The first set of approaches we will consider attempt to assign the weights by achieving an optimality condition using only information contained in the expert distributions themselves, without reference to any external information such as observed data.

Maximising entropy

In a context of near complete uncertainty about the relative reliabilities of the experts (information sources) it may be desirable to combine the prior distributions such that $\pi(\theta)$ is maximally diffuse. According to its proponents, such an approach would ensure that, given the constraints imposed by \mathbf{F}_θ , the pooled distribution is the one which best represents the current state of knowledge (Jaynes, 1957; Savchuk and Martz, 1994). In order to choose $\boldsymbol{\alpha}$ so as to maximise prior diffuseness, one can maximise the entropy of the log-pooled prior, i.e.

$$\begin{aligned} H_\pi(\theta; \boldsymbol{\alpha}) &= \mathbb{E}_\pi [-\log \pi(\theta)] = - \int_{\Theta} \pi(\theta) \log \pi(\theta) d\theta, \\ &= -\log t(\boldsymbol{\alpha}) - \sum_{i=0}^K \alpha_i \mathbb{E}_\pi [\log f_i]. \end{aligned} \tag{3.1}$$

Formally, we want to find $\hat{\alpha}$ such that

$$\hat{\alpha} := \arg \max_{\alpha} H_{\pi}(\theta; \alpha). \quad (3.2)$$

This approach, however, does not result in a convex optimisation problem, therefore one is not guaranteed to find a unique solution – see Remark 3.1 for intuition as to why. A possible resolution to the non-uniqueness of the maximum entropy solution would be to add further constraints, for instance requiring that $\mathbb{E}_{\pi}[\theta] = m$. It is however unclear which set of constraints would ensure uniqueness.

Minimising KL divergence to a reference measure

Let \mathbf{y} denote observed data and π^J denote the Jeffreys’s prior for θ . This is seen as a non-informative or ‘objective’ choice of prior distribution, and leads to a posterior distribution, $p^J(\theta | \mathbf{y})$, which is set to be dominated by the likelihood (Berger, 2006). Based on initial work by Bousquet (2008), Rufo et al. (2012a) propose assigning the weights in such a way that the Kullback-Leibler divergence between the Jeffreys’s posterior and the logarithmic pool is minimised:

$$\begin{aligned} L_R(\alpha, \mathbf{y}) &= \text{KL}(p^J(\cdot | \mathbf{y}) || \pi), \\ &= -H_{p^J} - \sum_{i=0}^K \alpha_i \mathbb{E}_{p^J} [\log f_i]. \end{aligned} \quad (3.3)$$

The main idea is to pick the weights so as to minimise prior-data conflict. However, the expression in (3.3) depends on observed data. Thus, Rufo et al. (2012a) propose computing the expected loss with respect to the marginal distribution of the data, $m(\mathbf{y}) := \int_{\Theta} L(\mathbf{y} | \theta) \pi^J(\theta) d\theta$, $L_R(\alpha) = \mathbb{E}_m [L_R(\alpha, \mathbf{y})]$. Now one can solve the optimisation problem

$$\begin{aligned} \hat{\alpha}_R &:= \arg \min_{\alpha} L_R(\alpha), \\ &= \arg \max_{\alpha} \left\{ \log t(\alpha) + \sum_{i=0}^K \alpha_i \mathbb{E}_m [\mathbb{E}_{p^J} [\log f_i]] \right\}, \end{aligned} \quad (3.4)$$

which Rufo et al. (2012a) show to lead to a unique pooled distribution under mild conditions – see below. Notice, however, that there could exist two vectors α_1 and α_2 such that $\pi(\theta | \alpha_1) \equiv \pi(\theta | \alpha_2)$ up to a set of null measure. Hence, the pooled **distribution** is unique, but the optimal weights might not be.

Minimising KL divergence between each opinion and the pool

One could also wish to choose the pooling weights so as to minimise the total Kullback-Leibler divergence between the pooled distribution, π , and each distribution in \mathbf{F}_{θ} . We can define a loss function $L(\alpha) = \sum_{i=0}^K \text{KL}(f_i || \pi)$ such that

$$L(\alpha) = -(K+1) \log t(\alpha) - (K+1) \sum_{i=0}^K \alpha_i \mathbb{E}_i [\log f_i] + \sum_{i=0}^K \mathbb{E}_i [\log f_i], \quad (3.5)$$

and we want to find

$$\hat{\boldsymbol{\alpha}} := \arg \min_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}). \quad (3.6)$$

Fortunately, this set up leads to a unique pooled distribution, a result we summarise in Remark 3.1.

Remark 3.1 (Uniqueness of the minimum KL solution). The distribution obtained following (3.6) is unique, i.e., there is only one aggregated prior $\pi(\theta | \boldsymbol{\alpha})$ that minimizes $L(\boldsymbol{\alpha})$.

Proof. See Appendix A (Carvalho et al., 2022). \square

By contrast, the problem in (3.2) requires one to minimise $\ln t(\boldsymbol{\alpha})$, hence lacking a sufficient condition for the existence of a unique solution. Likewise, using the loss function $L'(\boldsymbol{\alpha}) = \sum_{i=0}^K \text{KL}(\pi || f_i)$ would not lead to a unique solution. See commentary about (3.4) above and Appendix B (Carvalho et al., 2022) for implementation details. The choice of summing the KL divergences is ultimately arbitrary, but has the advantage of preserving the convexity of the optimisation problem and allowing efficient algorithms to be employed.

3.2 Hierarchical modelling of the weights

As discussed by Poole and Raftery (2000) and others (Zhong et al., 2015; Li et al., 2017), estimating the weights would be of interest since this would allow one to assess the reliability of each source of information (expert). Li et al. (2017) explore the idea of computing the pooled distribution for several values of the weights. Whilst informative, this approach has two issues: (a) it does not scale well with increasing the number K of distributions being combined, and; (b) it fails to account for any (posterior) dependence between model parameters and the weights. In this section we propose placing a hierarchical prior on the weights, allowing for standard Bayesian inference about these quantities.

A natural choice for a prior distribution for $\boldsymbol{\alpha}$ is the $(K + 1)$ -dimensional Dirichlet distribution

$$\pi_A(\boldsymbol{\alpha}) = \frac{1}{\mathcal{B}(X)} \prod_{i=0}^K \alpha_i^{x_i - 1}, \quad (3.7)$$

where $X = \{x_0, x_1, \dots, x_K\}$ is the vector of hyperparameters for the Dirichlet prior and $\mathcal{B}(\cdot)$ is the multinomial Beta function. The Dirichlet offers a simple, albeit potentially inflexible prior.

A more flexible prior for $\boldsymbol{\alpha}$ is the logistic-normal distribution (Aitchison and Shen, 1980):

$$\pi_A(\boldsymbol{\alpha} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}} \frac{1}{\prod_{i=0}^K \alpha_i} \exp \left(\left(\log \left(\frac{\boldsymbol{\alpha}_{-K}}{\alpha_K} \right) - \boldsymbol{\mu} \right)^\top \boldsymbol{\Sigma}^{-1} \left(\log \left(\frac{\boldsymbol{\alpha}_{-K}}{\alpha_K} \right) - \boldsymbol{\mu} \right) \right), \quad (3.8)$$

where α_{-K} represents the vector α without the K -th element, $\boldsymbol{\mu}$ is a K -size mean vector, and $\boldsymbol{\Sigma}$ is a $K \times K$ covariance matrix. (Aitchison and Shen, 1980) propose choosing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ minimizing the KL divergence between the Dirichlet (3.7) and the logistic-normal (3.8) distributions, i.e.

$$\begin{aligned}\mu_i &= \psi(x_i) - \psi(x_K), \quad i = 0, 1, \dots, K-1, \\ \Sigma_{ii} &= \psi'(x_i) + \psi'(x_K), \quad i = 0, 1, \dots, K-1, \\ \Sigma_{ij} &= \psi'(x_K),\end{aligned}$$

where $\psi(\cdot)$ is the digamma function, and $\psi'(\cdot)$ is the trigamma function.

The marginal prior for θ ,

$$\tilde{\pi}(\theta) = \int_{\mathcal{A}} \pi(\theta | \boldsymbol{\alpha}) \pi_{\mathcal{A}}(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \quad (3.9)$$

can also be efficiently approximated through Monte Carlo sampling when π can be written in closed-form. Even when it cannot be expressed analytically, it is still possible to sample from the marginal prior by using quadrature-based methods for computing $t(\boldsymbol{\alpha})$ when θ is unidimensional (see Discussion).

Concerning posterior inference, the marginal posterior for θ can be obtained through standard methods and shall not be discussed further. The next object to consider is the marginal posterior for the weights, $p(\boldsymbol{\alpha} | \mathbf{y})$, which can be obtained through

$$\begin{aligned}p(\boldsymbol{\alpha} | \mathbf{y}) &= \int_{\Theta} p(\boldsymbol{\alpha}, \theta | \mathbf{y}) d\theta, \\ &= \frac{\pi_{\mathcal{A}}(\boldsymbol{\alpha})}{c(\mathbf{y})} \int_{\Theta} L(\mathbf{y} | \theta) \pi(\theta | \boldsymbol{\alpha}) d\theta, \\ &\propto \pi_{\mathcal{A}}(\boldsymbol{\alpha}) \kappa(\boldsymbol{\alpha}, \mathbf{y}),\end{aligned}$$

where $c(\mathbf{y}) := \int_{\mathcal{A}} \int_{\Theta} p(\boldsymbol{\alpha}, \theta | \mathbf{y}) d\theta d\boldsymbol{\alpha}$.

In some situations, in particular the conjugate situation discussed in Section 2.1 and exemplified in the Applications section below, it is possible to write down $\kappa(\boldsymbol{\alpha}, \mathbf{y})$ in closed-form. This is very convenient because the posterior expectation of the weights, $E_p[\boldsymbol{\alpha} | \mathbf{y}]$, becomes $E_{\pi_{\mathcal{A}}}[\boldsymbol{\alpha} \kappa(\boldsymbol{\alpha}, \mathbf{y})]$, i.e., the expectation of a known function with respect to the prior on the weights. This expectation can be easily and accurately approximated with simple Monte Carlo techniques rather than MCMC – see Section 4.2 for example applications.

4 Applications

In this section we shall present a wide range of applications for logarithmic pooling, from prior elicitation to meta-analysis to Bayesian melding. Computational details, along with instructions to get reproducible code, are given in Appendix B (Carvalho et al., 2022).

4.1 Elicitation: combining expert priors on survival probabilities

The first example we consider is combining expert opinions about probabilities and proportions. We analyse an example proposed by Savchuk and Martz (1994) (also discussed in Rufo et al., 2012b) in which four experts are required to supply prior information about the survival probability θ of a certain unit. The experts express their opinion as prior means for the survival probability, which Savchuk and Martz (1994) then use to construct prior distributions with maximum entropy given the restriction on the means. From the vector of prior means $\mathbf{m} = \{m_0 = 0.95, m_1 = 0.80, m_2 = 0.90, m_3 = 0.70\}$, the authors obtain the parameters of the Beta distributions for each expert, $\mathbf{a} = \{a_0 = 18.10, a_1 = 3.44, a_2 = 8.32, a_3 = 1.98\}$ and $\mathbf{b} = \{b_0 = 0.955, b_1 = 0.860, b_2 = 0.924, b_3 = 0.848\}$. Furthermore, an experiment is conducted and $y = 9$ successes out of $n = 10$ trials are observed. Thus, in this application we are able to estimate the posterior distribution for the survival probability and also, with the hierarchical modelling approach, the posterior distribution for the weights in face of the observed data. We consider two hierarchical priors: a Dirichlet(1/10, 1/10, 1/10, 1/10) and a moment-matching logistic-normal prior (see Section 4.2 for justification).

The probability distribution of the survival probability for the i -th expert is a Beta distribution with (hyper)parameters a_i and b_i . The log-pooled distribution for θ is then

$$\begin{aligned} \pi(\theta) &\propto \prod_{i=0}^K f_i(\theta; a_i, b_i)^{\alpha_i}, \\ &\propto \theta^{a^* - 1} (1 - \theta)^{b^* - 1}, \end{aligned} \quad (4.1)$$

with $a^* = \sum_{i=0}^K \alpha_i a_i$ and $b^* = \sum_{i=0}^K \alpha_i b_i$. Note that (4.1) is the kernel of a Beta distribution with parameters a^* and b^* . Hence the entropy is the following

$$H_\pi(\theta) = \log \mathcal{B}(a^*, b^*) - (a^* - 1)\psi(a^*) - (b^* - 1)\psi(b^*) + (a^* + b^* - 2)\psi(a^* + b^*). \quad (4.2)$$

And the KL divergence between $\pi(\theta)$ and $f_i(\theta)$ is

$$\begin{aligned} d_i = KL(f_i || \pi) &= \ln \left(\frac{\mathcal{B}(a^*, b^*)}{\mathcal{B}(a_i, b_i)} \right) + (a_i - a^*)\psi(a_i) + (b_i - b^*)\psi(b_i) \\ &\quad - (a_i - a^* + b_i - b^*)\psi(a_i + b_i). \end{aligned} \quad (4.3)$$

In this conjugate setting, the posteriors associated with each expert are also Beta distributions with parameters $a'_i = a_i + y$ and $b'_i = b_i + (n - y)$. This allows us to employ the maximum entropy and minimum KL procedures to combine these posterior distributions and thus make the weights comparable with the posterior means obtained with the hierarchical priors.

Our analysis of this example is thus split into two: weights for the priors and for the posteriors. Before observing any data, we can employ the optimisation procedures discussed above to obtain weights only taking into account information encoded in the expert priors themselves. To these optimisation procedures we add the technique of Rufo

Method		α_0	α_1	α_2	α_3
Prior	Maximum entropy	0.00	0.00	0.00	1.00
	Minimum KL ¹	0.04	0.96	0.00	0.00
	Rufo et al. (2012a)	0.00	0.00	0.00	1.00
Posterior	Maximum entropy	0.00	0.00	0.00	1.00
	Minimum KL	0.17	0.83	0.00	0.00
	Dirichlet ²	0.26	0.24	0.27	0.23
	Logistic-normal ²	0.27	0.24	0.31	0.18
	Marginal likelihoods (BMA)	0.27	0.24	0.30	0.19

Table 1: **Weights obtained using different methods for the survival probability example (Savchuk and Martz, 1994).** ¹ – Kullback-Leibler; ² – Posterior mean for α .

et al. (2012b) which seeks to minimise KL distance between the pooled prior and the Jeffreys’s posterior. When data are available, we can then use maximum entropy and minimum KL to obtain the weights in the same fashion as before, but now also estimate the posterior distribution of weights using a hierarchical prior. Finally, for this example we can also compute the integrated (marginal) likelihood of each expert, meaning that we can, assuming one of the experts is correct, compute “model” probabilities by normalising the marginal likelihoods (see Section 4.2, below).

In Table 1 we present weights obtained with the optimisation methods for the priors, including the solution found by Rufo et al. (2012b) (Section 5.2 therein). With regard to the posteriors, we show maximum entropy, minimum KL along with posterior means of the weights under two prior distributions (Dirichlet and logistic-normal). Maximising the entropy of the pooled prior – and posterior – led to the degenerate solution $\alpha = \{0, 0, 0, 1\}$, which gives all the weight to the most diffuse prior distribution – Beta(1.98, 0.848). Since $t(\alpha)$ is concave, we expect to find the maximum entropy given by the boundary conditions, which may lead to points at the frontier of the simplex. Unsurprisingly, the same solution was found by Rufo et al. (2012b), whose method tends to favour more diffuse distributions. Minimising Kullback-Leibler divergence between the pooled prior and each expert prior leads to a unique solution but in this case also suggests to discard two of the opinions. The hierarchical priors gave very similar posterior distributions for the weights, which assign the experts nearly equal weight, although the logistic-normal prior led to results closer to Bayesian model averaging (BMA) weights, based on the normalised marginal likelihoods of each expert. Please see Section 4.2 below for how relying only on the posterior means can be misleading, however.

Table 2 contains the prior and posterior mean and credibility intervals from each of the methods and also the case in which we assign an equal weight ($1/K$) to each opinion. The prior densities for each expert and pooling method are show in Figure S4. Assigning equal weights actually gives a prior mean that is the same as the maximum likelihood estimate of θ , $\hat{\theta} = 9/10$. This explains why both hierarchical posteriors resemble equal weights so closely. Finally, we use the integrated (marginal) likelihood (Raftery et al., 2007, eq. 9), $l(y, n) = \int_0^1 f(y|\theta)\pi(\theta) d\theta$, as a univariate summary to compare the priors.

Method	Prior	Posterior
Equal weights	0.90 (0.64–1.00)	0.90 (0.73–0.99)
Maximum entropy	0.70 (0.17–0.99)	0.86 (0.63–0.98)
Minimum KL	0.82 (0.42–1.00)	0.87 (0.67–0.99)
Rufo et al. (2012b)	0.70 (0.17–0.99)	0.86 (0.63–0.98)
Dirichlet	0.86 (0.40–1.00)	0.89 (0.70–0.99)
Logistic-normal	0.88 (0.35–1.00)	0.89 (0.71–0.99)

Table 2: **Prior and posterior mean and credibility intervals for each method for assigning the weight, survival probability example (Savchuk and Martz, 1994).** Values for the hierarchical priors are from the marginal prior of θ in (3.9).

Expert priors		Pooled priors	
Expert 0	0.237	Equal weights	0.254
Expert 1	0.211	Maximum entropy	0.163
Expert 2	0.256	Minimum KL	0.223
Expert 3	0.163	Hierarchical prior ¹ (Dirichlet/logistic-normal)	0.255

Table 3: **Integrated likelihoods for the priors of each expert as well as the combined priors, failure probability example.** For the hierarchical priors we take the posterior expectations of a^* and b^* as a_i and b_i , respectively. ¹ – Calculated using the posterior mean of α .

The marginal likelihood for the i -th expert is $l(y, n \mid a_i, b_i)$ and its form is given in any elementary textbook on Bayesian statistics. For the hierarchical priors we take the posterior mean of (a^*, b^*) as (a_i, b_i) . Results are given in Table 3 and show that, apart from expert 3 – and hence the maximum entropy pooled prior –, all other pooled priors and individual experts' priors give similar marginal likelihoods.

The BMA weights are $\alpha'' = \{0.27, 0.24, 0.30, 0.19\}$, which are not very different from equal weights. In Table 1 we see that the posterior distribution for the weights estimated under both priors favours expert 2, the expert with the highest marginal likelihood. The logistic-normal gives expert 2 a higher weight when compared with the Dirichlet. This is connected to the increased flexibility of the logistic-normal (see Section 4.2).

We stress that the marginal likelihoods are not being used here as a means of selecting priors, but rather as a useful univariate summary which is informative about the compatibility with the observed data and hence informative about prior-data conflict. While in this example one can gain insight into prior-data conflict from just the prior means and y/n , in other situations it might be harder to discern which expert gave the best (prior) guess.

4.2 Posterior distribution of the weights: interpretability and prior sensitivity

A few natural questions arising from the analyses presented in the previous section are: how can one interpret the posterior distribution over α in light of data? How does the

prior on the weights affect inferences? In this section we provide a few experiments to elucidate these questions.

Induced distribution on θ

We begin our investigation by studying the effect of the choice of $\pi_A(\boldsymbol{\alpha})$ on the conditional densities for the parameter of interest, $\pi(\theta \mid \boldsymbol{\alpha})$, and also the marginal prior $\tilde{\pi}$ (3.9). In particular, we are interested in understanding how the expert opinion configurations \mathbf{F}_θ interact with the prior on the weights. For concreteness, consider a situation with $K = 3$ experts who all elicit Beta distributions about $\theta \in (0, 1)$. In particular, we have $\mathbf{a} = \{22.4, 12.0, 1.6\}$ and $\mathbf{b} = \{201.6, 12, 0.18\}$ such that the prior expectations according to each expert are 0.1, 0.5 and 0.9, respectively. Moreover, all distributions have the same coefficient of variation, $\sqrt{\text{Var}(\theta)}/E[\theta] = 0.2$. Throughout we will consider four prior distributions on $\boldsymbol{\alpha}$: a Dirichlet(1, 1, 1), written Dirichlet(1) for short and its corresponding moment-matching logistic-Normal distribution; and a Dirichlet(1/10) and its moment-matching logistic-Normal.

Now we can ask what shapes for $\pi(\theta \mid \boldsymbol{\alpha})$ are induced by the prior on π_A . Figure 1 shows the conditional pooled prior densities and we can see that the Dirichlet(1) and Logistic-normal(1) priors lead to induced conditional priors that skew towards the highest density expert (expert 0 here) and this leads to a marginal prior with larger mass around 0.1 (see marginal prior in Figure S5). Moreover, the Logistic-normal prior seems to be slightly more flexible in that it allows a few shapes away from expert 0. The Dirichlet(1/10) and Logistic-normal(1/10) priors on $\boldsymbol{\alpha}$ on the other hand lead to conditional densities that encompass all three experts, and this manifests marginally as a multimodal prior on θ . As we will see, these different behaviours will have implications for the ability to learn and interpret the weights.

Generative modelling and calibration

One can now ask questions about the ability to learn the weights in the logarithmic pool. We start by considering a generative approach: if one generates data according to the hierarchical model implied by the prior structure discussed so far, is it possible to recover the generating weights?

Consider the generative model:

$$\begin{aligned}\boldsymbol{\alpha} &\sim \Pi_A(X), \\ \theta \mid \boldsymbol{\alpha} &\sim \Pi(\cdot \mid \boldsymbol{\alpha}), \\ \mathbf{y} \mid \theta &\sim F(\cdot \mid \theta).\end{aligned}\tag{4.4}$$

One can use this model to generate data \mathbf{y} and then compute $p(\theta, \boldsymbol{\alpha} \mid \mathbf{y})$ in order to assess whether the model is well-calibrated in the sense of Talts et al. (2018). The main idea is to sample N times from the generative model, generating N triples $(\boldsymbol{\alpha}^{(n)}, \theta^{(n)}, \mathbf{y}^{(n)})$, $n = 1, \dots, N$. One can then draw L (possibly approximate) samples from $p(\theta, \boldsymbol{\alpha} \mid \mathbf{y}^{(n)})$. If the model is well-calibrated with respect to the sampling method, the ranks of $(\theta^{(n)}, \boldsymbol{\alpha}^{(n)})$

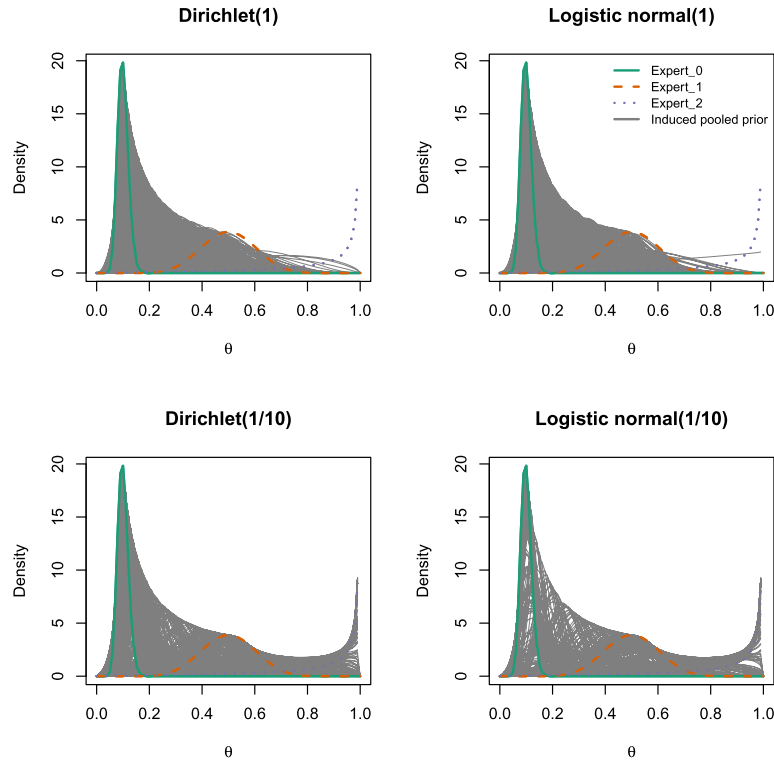


Figure 1: **Conditional induced prior densities under different choices of the prior on the weights.** We show the three expert opinions along with the induced density $\pi(\theta \mid \alpha)$ for 1000 simulations from the prior on the weights, π_A , for various choices of π_A .

with respect to the posterior should be uniformly distributed on $L + 1$. In particular, we will run 1000 simulations from the generative model, draw approximate samples from using Hamiltonian Monte Carlo (HMC, see Appendix B) and check whether the ranks of the true simulated values are uniform.

In Figure 2 we present the empirical cumulative distribution function (ECDF) of the observed ranks for all parameters in the model, including the weights (α), the success probability (θ) and the induced Beta hyperparameters (a^* and b^*). The ECDFs of the ranks are presented alongside 95% confidence intervals for the theoretical CDF. The results show that ECDF falls within its theoretical uncertainty bounds (blue shaded bands), indicating that the model is well-calibrated, and one can be confident that the generative model in (4.4) leads to an inferentially tractable model, at least for the particular choice of hyperparameters taken here.

While calibration is desirable, one might also want to study how the posterior distribution for α behaves as the evidence and its strength change. To make this concrete,

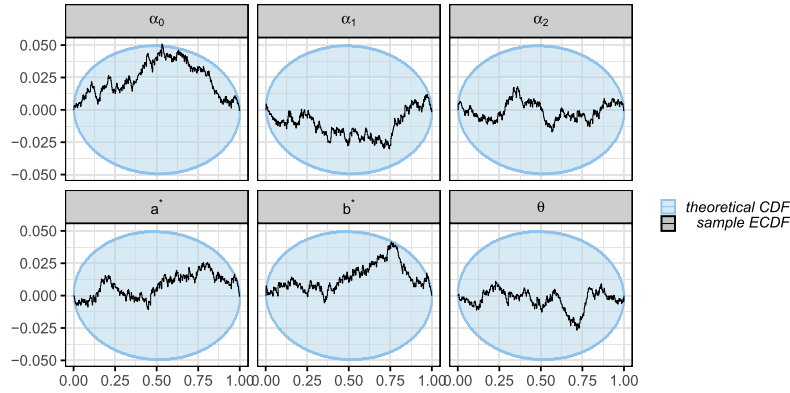


Figure 2: **Empirical cumulative distribution of ranks in simulation-based calibration of the logarithmic pooling model.** We show the empirical CDF (black line) along with its 95% confidence band (blue). One can claim that the model and inference apparatus are well-calibrated if the black line stays within the blue shaded area, i.e., within the confidence bands for all parameters. Here we have simulated 1000 data sets from the generative model in (4.4) with $X = (1/10, 1/10, 1/10)$ and $n = 10$ – see text for the experts’s hyperparameters.

consider the situation where one has the same Beta distributions as described above, with prior means 0.1, 0.5 and 0.9, respectively. Now one observes $y = 4$ successes in n trials, leading to a summary statistic of $\bar{y}_n = 0.5$. We can then ask what the (marginal) posterior for α would look like. What if we observed $y = 400$ in $n = 1000$ trials? Figure 3 shows the posterior mean and 95% credibility intervals for each component of α as both \bar{y}_n and n vary, under the Dirichlet(1/10) and Logistic-normal(1/10) priors, along with BMA weights.

We see that the posterior means resemble the BMA weights, but there is substantial uncertainty regarding the weights even when n is very large. We also see that the posterior means display a “shrinkage” effect whereby they never quite match the BMA weights exactly, an effect even more salient when the Dirichlet(1) and Logistic-normal(1) priors are employed – see Figure S6. These results are not surprising given those in Figure 1: when the prior on the weights leads to conditional priors that do not span the whole space of expert priors, one cannot hope to obtain appreciable posterior weight for a given expert, even in the presence of overwhelming data.

In Figure 4 we show the full joint posterior of the weights (α) for a situation where expert 1, whose posterior mean for θ is 0.5, should be favoured, i.e. the sample mean is $\bar{y}_n = 1/2$. We see that when one employs the Dirichlet(1/10) prior, one can usually achieve a posterior mean for the weights (solid circle) that is close to the BMA weights (solid triangle). Also, while for $n = 10$ Figure 4b shows samples in all three corners meaning that at least some posterior mass was concentrated on the corners, when $n = 10,000$ Figure 4d shows that the posterior concentrates around expert 1 as expected, albeit with substantial uncertainty as discussed above.

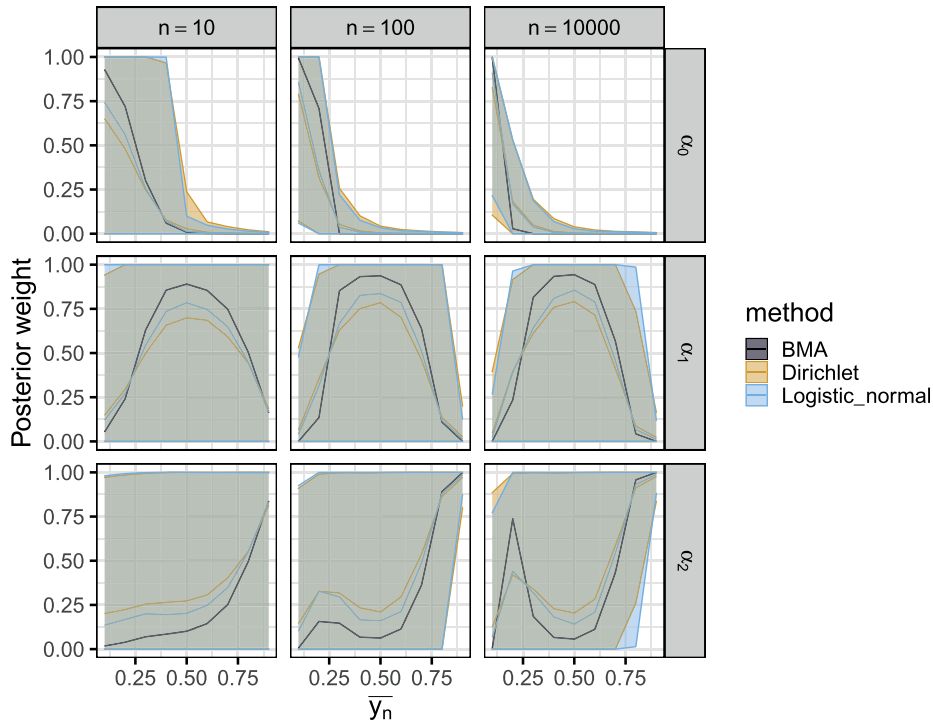


Figure 3: **Posterior distribution of the weights with varying strengths of evidence.** Here we show the posterior mean and 95% central BCIs for the weights in the three-expert scenario (see text) with the Dirichlet(1/10, 1/10, 1/10) prior and its moment-matching logistic-normal. We also show the Bayesian model averaging (BMA) weights for comparison (black line). The sample mean varies in the x-axis and vertical panels show different data set sizes.

4.3 Bayesian melding with varying weights

Another important application of logarithmic pooling is in the Bayesian melding method of Poole and Raftery (2000). Deterministic simulation models are widespread in Science and Engineering (see Poole and Raftery (2000) and references therein). One is often interested in a deterministic model M with inputs $\theta \in \Theta \subseteq \mathbb{R}^p$ and outputs $\phi \in \Phi \subseteq \mathbb{R}^q$, such that $\phi = M(\theta)$. If one wants to learn about θ from data and a (prior) distribution on ϕ is available, then one needs a method to combine the information between the prior on θ and the prior induced on it through M , which is often non-invertible.

Bayesian melding seeks to draw inference by first employing logarithmic pooling to construct a prior on ϕ of the form

$$\tilde{q}_{\Phi}(\phi) \propto q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha}, \quad (4.5)$$

where $q_1^*(\cdot)$ is the **induced** prior on the outputs and q_2 is the prior on ϕ without con-

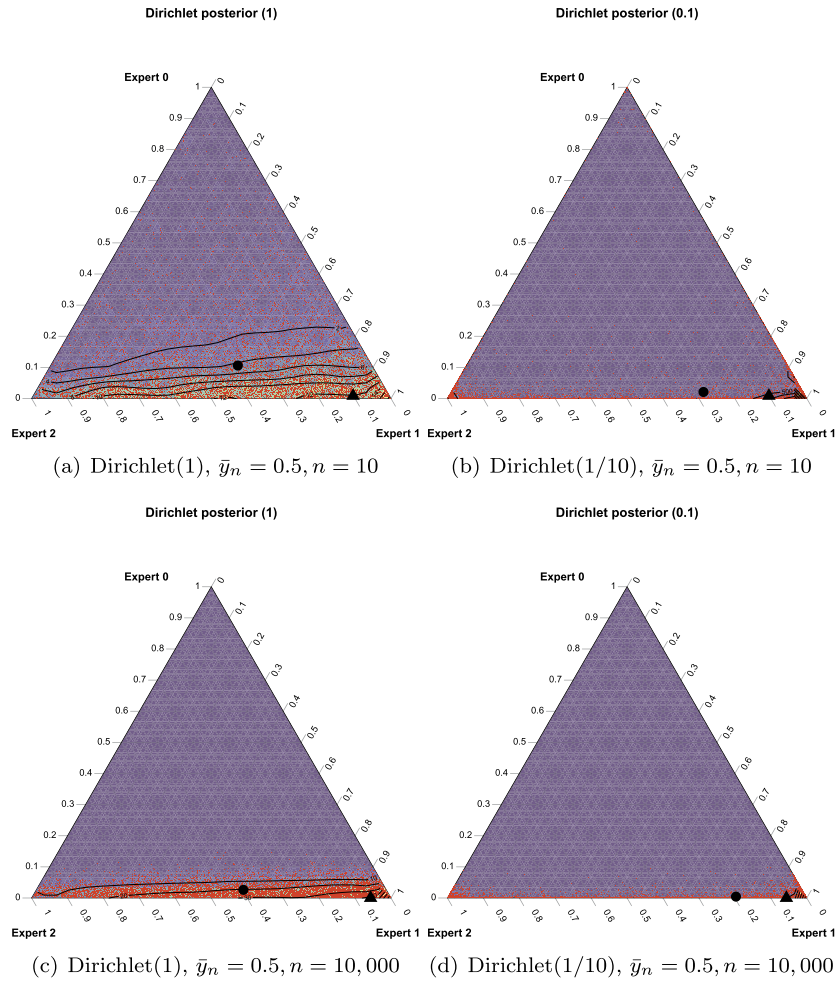


Figure 4: **Full joint posterior of the weights, Beta example.** We show the full posterior of α for a situation where the sample mean \bar{y}_n is 0.5, for $n = 10$ and $n = 10,000$ under different priors on the weights – see text for the expert hyperparameters. The solid circle marks the posterior mean and the solid triangle marks the BMA weights. Small red dots show the posterior samples and density contours are overlaid. Purple shows the regions of low posterior density, while yellow regions depict higher density. This figure was made with the **Ternary** package in R (Smith, 2017).

sidering the deterministic model, henceforth called the natural prior on ϕ . The prior in (4.5) can then be inverted to obtain a *coherised* prior on θ , $\tilde{q}_\Theta(\theta)$. Poole and Raftery (2000) give a way of obtaining \tilde{q}_Θ even when M is non-invertible, which we will not discuss further here – see Section 3.4 in Poole and Raftery (2000).

Standard Bayesian inference may then follow, leading to the posterior

$$p_{\Theta}(\theta) \propto \tilde{q}_{\Theta}(\theta)L_1(\theta)L_2(M(\theta)), \quad (4.6)$$

which enjoys all the properties of usual posterior distributions. The method allows standard Bayesian inference to be carried out about all quantities of interest in the model, which makes it attractive to application in policy making (Alkema et al., 2008), where proper acknowledgment of uncertainty is crucial.

In Poole and Raftery (2000) (Section 6.2 therein), the authors fix $\alpha = 1/2$, justifying their choice by the fact that while the weights should reflect the reliability of each expert (information source), in the specific context of Bayesian melding one is combining distributions based on different bodies of evidence, but assessed by the same expert. Another option is to fix $\alpha = 1 - \epsilon$, with ϵ small (Alkema et al., 2007). This can be useful when the prior distribution on outputs is uniform, as it still enforces the constraint, but keeps the prior information about the inputs. Here we relax the restriction of fixing the weight, instead modelling α through a hyperprior.

We now turn our attention to applications of logarithmic pooling to the statistical analysis of deterministic models. In their seminal paper, Poole and Raftery (2000) lay out Bayesian melding as a way to achieve full Bayesian inference for deterministic models. In this section we explore two Bayesian melding applications and extend their approach by accommodating uncertainty about the weight α .

Bowhead whale population growth

We begin with the analysis of a non-age-structured population deterministic model (PDM) population model for bowhead whales originally carried out by Poole and Raftery (2000). The model describes the annual population of bowhead whales in terms of the annual number of whales killed, C_t , the maximum sustainable yield rate (MSYR) and the initial bowhead population (P_0) as:

$$P_{t+1} = P_t - C_t \times \text{MSYR} \times P_t (1 - (P_t/P_0)^2). \quad (4.7)$$

One of the quantities of interest in the model was P_{1993} , due to 1993 being the last year for which independent abundance measurements were available, allowing for model calibration. Another important model quantity is the rate of population increase from 1978 to 1993, ROI, defined through

$$P_{1993} = P_{1978}(1 + \text{ROI})^{15}.$$

We are then interested in the model outputs $\phi = \{P_{1993}, \text{ROI}\}$. The key idea is to account for the influence of the priors on the inputs $\theta = \{\text{MSYR}, P_0\}$ on P_{1993} through the **induced** distribution. In particular, we aim at composing the prior distribution

$$\tilde{q}_{\Phi}(P_{1993}) \propto q_1^*(P_{1993})^{\alpha} q_2(P_{1993})^{1-\alpha}, \quad (4.8)$$

where q_1^* is the induced distribution and q_2 is the natural prior on P_{1993} . The main innovation we propose here is to place a probability distribution over α in order to relax

the need to fix it to particular value. We choose a Beta(1, 1) prior as our π_A . The target posterior is then

$$p_{\Theta, M}(P_0, \text{MSYR}, \alpha \mid C_t) \propto \tilde{q}_{\Theta}(P_0, \text{MSYR})L_1(P_0, \text{MSYR})L_2(P_{1993})\pi_A(\alpha), \quad (4.9)$$

where \tilde{q}_{Θ} is the suitably inverted distribution over the input space from the prior over the output space, \tilde{q}_{Φ} (see Poole and Raftery (2000), section 3.3.4). The subscript makes reference to the fact that this is a posterior over the inputs $\theta \in \Theta$ which are linked to the outputs $\phi \in \Phi$ by a deterministic model M , given by (4.7). Further details on priors and likelihoods are given in Poole and Raftery (2000) and Appendix D (Carvalho et al., 2022) of this paper. We note that when α is random, it is important to include all of the normalising constants that depend on it (Neuenschwander et al., 2009), in particular the normalising constant of the expression in (4.8).

Here we will consider two ways of approximating (4.9). First, we used the sampling importance-resampling (SpIR) algorithm described in Appendix B. This method does not rely on any parametric approximation to the induced distribution q_1^* , instead using standard kernel methods to approximate the density at any point. We used $k = l = 100,000$ iterations to produce a sample from $p_{\Theta, M}$. We also explored a Hamiltonian Monte Carlo (HMC) implementation in Stan (Carpenter et al., 2017). However, for this implementation we needed to approximate q_1^* by a parametric form. Since q_2 is a normal distribution, we approximate q_1^* by a normal distribution such that \tilde{q}_{Φ} (4.8) can be written in closed-form. We give further discussion on this choice in Appendix D (Carvalho et al., 2022). Since $p_{\Theta, M}$ is a challenging target distribution, we used four independent chains of 10,000 iterations each. We observed a low percentage of divergent iterations (<2%), likely caused by the very challenging posterior geometry induced by high correlations between parameters.

In Figure 5 we show the marginal posteriors for various quantities of interest, obtained with both algorithms and for fixed and varying α . As expected, SpIR are a bit noisier, but distributions are largely the same as obtained by MCMC. For α in particular, despite the ruggedness of distribution obtained with SpIR, the mean and 95% credibility intervals of both distributions match very closely: $\hat{\alpha}_{\text{SpIR}} = 0.39$ (0.02–0.87) and $\hat{\alpha}_{\text{MCMC}} = 0.40$ (0.02–0.91). The high posterior uncertainty about α and the substantial overlap between distributions with fixed and varying α could be explained by the lack of sensitivity of the posterior distribution to α . We confirm this is indeed the case by running SpIR (original algorithm by Poole and Raftery (2000)) for a few values of α (including the endpoints 0 and 1) and verifying very little difference in the resulting posteriors (Figure S3, Carvalho et al., 2022).

Influenza in a boarding school

Another important class of deterministic models are the ordinary differential equation-based models of disease transmission. Here we will consider such a deterministic epidemic model and how one can draw inference about a key epidemiological quantity, the basic reproductive number, R_0 . In 1978, an anonymous source reported an influenza H1N1 epidemic at a small boarding school in England (Anon., 1978). In total, 512 boys

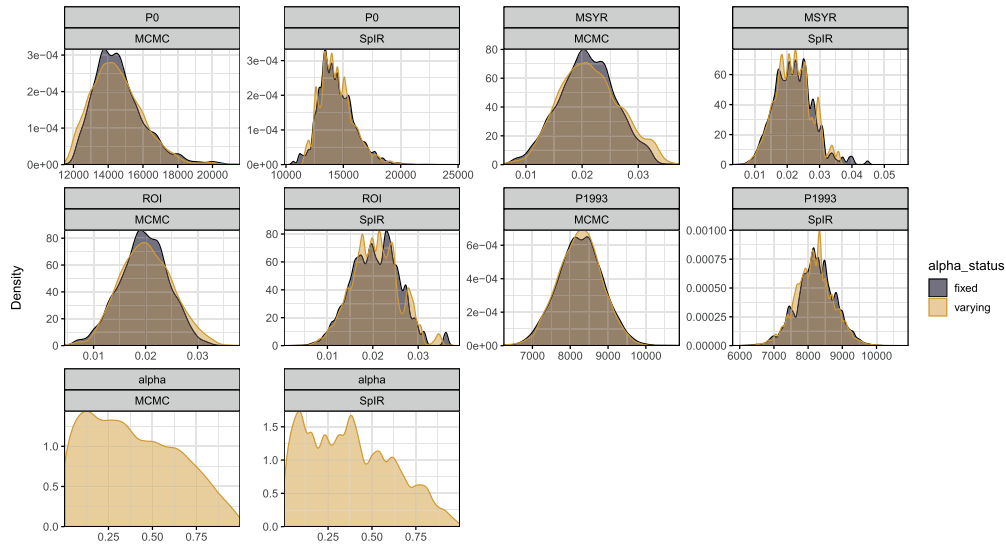


Figure 5: **Marginal posterior distributions for various quantities of interest in the bowhead population model.** We show the posterior distributions obtained by using sampling importance-resampling (SpiR) and Markov chain Monte Carlo (HMC-MCMC), for fixed $\alpha = 1/2$ and placing a prior π_A on α (“varying”).

out of 763 became ill during the outbreak. Due to the population being isolated and having high rates of contact, many of the assumptions of compartmental epidemic models hold. In particular, the Susceptible-Infected-Removed (SIR) model is a good description of disease spread. The model consists of the system of ordinary differential equations

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI, \\ \frac{dI}{dt} &= \beta SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}$$

where $S(t) + I(t) + R(t) = 1 \forall t$, β is the transmission (infection) rate and γ is the recovery rate. The basic reproductive number is $R_0 = \frac{\beta}{\gamma}$ and the goal is to draw inference about β and γ , and consequently about R_0 , from data. Data on the number of infected individuals per time ($Y(t)$) were obtained from the **outbreaks** package (Jombart et al., 2019) and we choose to model the deviation from the ODE solution using log-normal errors, i.e.,

$$L(Y(t) \mid \beta, \gamma, \sigma_I^2) = \text{log-normal}(\mu = \log(I(t)), \sigma_I^2), \quad (4.10)$$

where $I(t)$ is computed *via* an ODE solver. Here we will consider a situation where one has priors on β and γ , which induce a prior q_1^* on R_0 , and also a prior q_2 on R_0 directly. This is the case when, for instance, one wants to make q_2 informative so as to

incorporate expert knowledge and/or evidence from previous study. For the priors on β and α we choose commonly used, so-called “uninformative” log-normal priors with parameters $\mu_\beta = \mu_\gamma = 0$ and $\sigma_\beta^2 = \sigma_\gamma^2 = 1$, which induces a log-normal distribution (q_1^*) on R_0 with parameters $\mu_1 = \mu_\beta - \mu_\gamma$ and $\sigma_1^2 = \sigma_\beta^2 + \sigma_\gamma^2$. Using the extensive information gathered by Biggerstaff et al. (2014), we constructed an informative log-normal prior (q_2) with mean 1.5 and variance 0.25^2 , which gives $\mu_2 = 0.3917656$ and variance $\sigma_2^2 = 0.1655264$. This leads to a prior credibility interval of (1.070–2.047), which covers most of the estimates (and confidence intervals) of R_0 for Influenza found by Biggerstaff et al. (2014). The target posterior is then

$$p(\beta, \gamma, \alpha | Y(t)) \propto L(Y(t) | \beta, \gamma, \sigma_I^2) q_1^*(R_0)^\alpha q_2(R_0)^{1-\alpha} \pi_A(\alpha), \quad (4.11)$$

where we again let π_A be a Beta(1,1) distribution. This setup is convenient because it leads to a closed-form expression for the combined prior on R_0 (see Appendix E (Carvalho et al., 2022)), while the log-normal priors are flexible and useful in practice. We approximate the posterior in (4.11) using HMC as described in Appendix B.

In Figure 6a we show the posterior distribution of the pooling weight α , which favours high values with a mean and 95% credibility interval of 0.77 (0.21–0.99). The posterior distribution for R_0 obtained by letting α vary and also the resulting distributions of fixing $\alpha = 1/2$ or $\alpha = 1$ are shown in Figure 6b. One can see that fixing $\alpha = 1$ and hence excluding the informative prior leads to a higher estimate of R_0 and fixing $\alpha = 1/2$ as per Poole and Raftery (2000) leads to the lowest estimates. The solution proposed in this paper, namely assigning α a prior and estimating it from data, leads to an intermediate solution. Fixing $\alpha = 1/2$ also leads to underestimating the measured incidence (Figure 6c), whilst setting $\alpha = 1$ leads to mean predictions that are higher, albeit still underestimating the measured incidence. Again, letting α vary leads to an intermediate solution.

Our results agree partially with the estimate obtained by Murray (2002), who finds $\rho = N/R_0 = 202$ and hence $R_0 = 3.78$, using purely numerical methods with no acknowledgment of uncertainty. The highest estimates we obtained were for fixed $\alpha = 1$, $R_0 = 3.02$ (2.27–3.83). This example showcases a desirable consequence of letting α vary: when the “natural” prior q_2 – which is normally informative – is incompatible with the data, it will receive a lower weight (α closer to 1) and hence allow the induced prior (q_1), which is usually more diffuse, to dominate. In fact, as discussed by Biggerstaff et al. (2014), the spread of the 1978 boarding school epidemic is unusually fast when compared to regular seasonal Influenza and was likely caused by the lack of previous exposure of the population to the causing strain, H1N1. The varying α approach makes it possible to deal with such an outlier data set by lowering the influence of the informative prior constructed based on previous studies. This is similar in spirit to ensemble forecasting (Leutbecher and Palmer, 2008), where by virtue of including many models one is able to achieve better forecasts than each individual model alone could.

5 Discussion

In this paper we have provided an overview of statistical applications of logarithmic pooling (LP), including a new approach based on assigning a prior measure to the

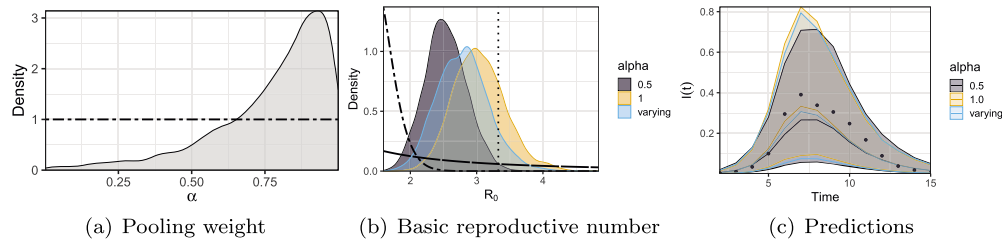


Figure 6: **Estimates of the pooling weight (α), the basic reproductive number (R_0) and predictions of the number of infected individuals.** The posterior distribution for the pooling weight α is shown in panel (a), where the horizontal dashed line shows the prior density, a Beta(1, 1). Panel (b) shows the posterior distribution for R_0 obtained with estimating α (“varying”) or fixing it to either 1/2 or 1. Vertical line shows $R_0 = 3.78$ (Murray, 2002), the dot-dashed line shows the informative prior q_2 and the longdash line shows the induced prior q_1^* . In panel (c) we show the posterior mean and 95% credibility intervals for the proportion of infected individuals, again by either letting α vary or fixing it to either 1/2 or 1.

weights. In what follows we discuss our findings in light of the rich literature on log-pooling, as well as point out connections to other parts of the statistical literature on model and forecast aggregation.

5.1 Objections to logarithmic pooling and their counter-arguments

West (1984) argues that LP is strictly theoretically justifiable only when the expert opinions agree. Moreover, LP also violates basic coherence in other respects, for instance when one considers marginalisation or other probability manipulations. Genest and Zidek (1986, p. 124) explain, however, that these conclusions stem from the restrictive assumption that the group utility is expressed as a function of the individual utilities. In a statistical application context, the expert opinions are usually employed by an independent decision maker, henceforth called the analyst, and she has her own utility function which can be assumed to not depend on the individual utilities.

A consequence of encoding opinions as probability densities is that representations of the same information might have different properties depending on the choice of dominating measure. The results of the meta analysis application presented in Appendix C (Carvalho et al., 2022) make this clear: choosing to represent the information brought by the studies as Beta distribution or a Gaussian does not affect the numerical values of means and probability intervals, but does seem to impact the optimality-based methods for choosing the weights, in particular minimum KL (Table 5). On the other hand, as shown by the agreement of the probability intervals in the bottom of Table 6, moving away from optimality criteria and instead assigning a prior distribution to the weights largely removes dependence on specific choices of probability densities by properly accommodating uncertainty about the weights.

One might worry about being able to learn the weights from data, since the weights depend on the likelihood only indirectly. Indeed, as shown in Section 4.2, it might not always be possible to identify the expert whose opinion is most consistent with the observed data given the uncertainty in the posterior distribution of the weights. Our experiments show that the posterior mean qualitatively resembles the Bayesian model averaging weights, even if one needs to be careful with the specification of $\pi_A(\boldsymbol{\alpha})$. As show in Section 3.2, the posterior means can be easily obtained via a simple Monte Carlo scheme in many situations, but we stress that it is important to take the whole posterior distribution into account (see Figure 4).

One might also be concerned with the long-run (frequentist) properties of the posterior for $\boldsymbol{\alpha}$. To understand what happens when the data set size grows, we ran a simple two-expert experiment using normal priors and a normal likelihood with known variance. Figure S7 shows that a Beta(1/10, 1/10) allows the posterior mean to concentrate on the “correct” expert as n grows. These results are interesting inasmuch as one would expect to lose the ability to learn the weights as the data size grows, owing to the fact that the weights are related to the data only through the prior, which becomes increasingly irrelevant to the posterior on the quantity of interest, θ . We posit that concentration around the “correct” expert will happen whenever the model is (a) well-specified and (b) regular. From (4.4) it is clear that one can think of a log-linear mixture as a hierarchical model. By being well-specified, we mean that the data are generated according to such a model. By regular we mean that the true generating parameters (including the weights) are in the interior of the support and other common regularity conditions. In this case, reasoning about posterior concentration would proceed analogously to what is done for hierarchical or “random effects” models – see Example 5 in Bochkina and Green (2014). The question of which classes of priors allow for fast concentration (or any concentration at all) is a very interesting avenue for further research.

As a final caveat, we note that if interest lies on a multivariate quantity $\theta \in \mathbb{R}^d$, $d \gg 1$, obtaining the normalising constant $t(\boldsymbol{\alpha})$ will entail computing a high-dimensional integral, which is infeasible to do via quadrature. Here, importance sampling techniques can be leveraged to provide stable and accurate estimates of normalising constants (see Future directions).

5.2 The case for (hierarchical) logarithmic pooling

We shall now argue that properties such as external Bayesianity, relative propensity consistency and log-concavity make logarithmic pooling a powerful tool for the analyst.

Mainly due to the simplicity of their construction, linear mixtures are much more popular in statistical applications (Frühwirth-Schnatter et al., 2019) than their log-linear cousins. As we hope to have shown in this paper, however, log-linear mixtures (logarithmic pooling) can be as useful or more. External Bayesianity means one does not need to worry about combining the priors first and then obtaining the posterior; one can simply take a set of posterior distributions computed with the same likelihood and combine them.

Moreover, LP preserves log-concavity, which might be crucial in computationally demanding settings where slice sampling, variational inference or other algorithms that assume log-concavity are employed. In summary, we argue that by employing logarithmic pooling to combine probability densities, the analyst is making the best use of the available information by forming a coherent distribution, that preserves many of the features encoded by the experts in their opinions.

After its theoretical properties, the strongest argument in favour of LP is by far its adaptability. The extra flexibility brought on by the hierarchical prior on the weights might prove crucial in scientific applications where decision under uncertainty is a regular occurrence. For example, a main strength of Bayesian melding is downweighting parameter values based on implausible model outputs. This strength is magnified by using a hierarchical prior that allows the weight parameter to vary. Indeed, Poole and Raftery (2000) (Section 5.2) argue that estimating α would be a fruitful path to explore and our results corroborate that view. The result in Section 4.3 makes clear the potential of varying-weights Bayesian melding for resolving prior-data conflict. In particular, for protecting the analyst from drawing strong conclusions when the “natural” prior on the quantity of interest is in disagreement with the information brought by the data under analysis.

When comparing the hierarchical prior approach to optimality-based procedures, one might argue that excluding a few or even all experts but one is not problematic since a few experts may, when suitably combined, summarise the information provided by the whole group. Whilst the weights are not probabilities, we argue that it would be preferable to have a solution that respects the so-called Cromwell’s rule (Lindley, 2013, p. 91), i.e., not assigning zero probability to events that are logically possible. Here this means allowing for the possibility that the opinion of all experts receives non-zero weight. Incidentally, this should also help alleviate some of the problems discussed in the previous section.

5.3 Future directions

Future research will explore further applications of logarithmic pooling in statistical learning such as combining several posterior predictive distributions from different models fitted to the same data. Techniques such as Bayesian predictive synthesis (BPS, McAlinn et al., 2018, 2019; McAlinn and West, 2019) and stacking (Yao et al., 2018) have focused on generalising linear pools to combine probabilistic predictions, and logarithmic pooling could be explored as possibility that preserves characteristics such as log-concavity and relative propensity consistency. BPS can include logarithmic pooling as a special case, but understanding the conditions under which this holds remains an open question.

Another interesting avenue for the future is studying the interaction between variable transformations and logarithmic pooling. An example is a situation where one has distributions about a probability p but is interested in the log-odds, $\omega = \log(p/(1-p))$. Should the experts be judged by how reasonable their distributions look in transformed space? How to assign the weights in this situation?

In a practical setting, one might have a collection of MCMC (approximate) samples from different posterior distributions. The statistical question then becomes how to sample from the pooled distribution by re-using these samples. Such task would likely necessitate specially-designed MCMC methods, and would constitute a rich area of future inquiry.

In such a setting, where both prior information and data are available, pooling methods are not the only option for evidence aggregation. A conspicuous example is Bayesian Evidence Synthesis (Jackson et al., 2015), a graphical model-based strategy which employs directed acyclic graphs (DAGs) and Bayes's rule to coherently update the posterior distribution of parameters. A comparison between these approaches in a specific applied example might constitute an interesting route of inquiry.

In closing, we hope this paper (i) showcases the usefulness – and potential pitfalls – of logarithmic pooling as a way of combining probability distributions and (ii) entices the statistical community to add it to their toolbox.

Supplementary Material

Supplementary Material of “Bayesian Inference for the Weights in Logarithmic Pooling” (DOI: [10.1214/22-BA1311SUPP](https://doi.org/10.1214/22-BA1311SUPP); .pdf). Supplementary Material contains:

Appendix A. Proofs.

Appendix B. Computational details: MCMC schema, sampling importance resampling with varying weights.

Appendix C. Meta-analysis of HIV prevalence studies in Brazil using log-pooling.

Appendix D. Details on the analysis of the Bowhead population growth model.

Appendix E. Pooling of common distributions: Poisson, Gamma, Log-normal.

Appendix F. Supplementary figures.

References

- Abbas, A. E. (2009). “A Kullback-Leibler view of linear and log-linear pools.” *Decision Analysis*, 6(1): 25–37. [224](#), [228](#)
- Aitchison, J. and Shen, S. M. (1980). “Logistic-normal distributions: Some properties and uses.” *Biometrika*, 67(2): 261–272. MR0581723. doi: <https://doi.org/10.2307/2335470>. [230](#), [231](#)
- Alkema, L., Raftery, A. E., and Brown, T. (2008). “Bayesian melding for estimating uncertainty in national HIV prevalence estimates.” *Sexually Transmitted Infections*, 84(Suppl 1): i11–i16. [240](#)

- Alkema, L., Raftery, A. E., Clark, S. J., et al. (2007). “Probabilistic projections of HIV prevalence using Bayesian melding.” *The Annals of Applied Statistics*, 1(1): 229–248. MR2393849. doi: <https://doi.org/10.1214/07-AOAS111>. 240
- Anon. (1978). “Influenza in a Boarding School.” *The British Medical Journal*, 1: 587. 241
- Bagnoli, M. and Bergstrom, T. (2005). “Log-concave probability and its applications.” *Economic Theory*, 26(2): 445–469. MR2213177. doi: <https://doi.org/10.1007/s00199-004-0514-4>. 226
- Berger, J. (2006). “The case for objective Bayesian analysis.” *Bayesian Analysis*, 1(3): 385–402. MR2221271. doi: <https://doi.org/10.1214/06-BA115>. 229
- Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., and Finelli, L. (2014). “Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature.” *BMC Infectious Diseases*, 14(1): 480. 243
- Bochkina, N. A. and Green, P. J. (2014). “The Bernstein–von Mises theorem and nonregular models.” *The Annals of Statistics*, 42(5): 1850–1878. MR3262470. doi: <https://doi.org/10.1214/14-AOS1239>. 245
- Bousquet, N. (2008). “Diagnostics of prior-data agreement in applied Bayesian analysis.” *Journal of Applied Statistics*, 35(9): 1011–1029. MR2522125. doi: <https://doi.org/10.1080/02664760802192981>. 229
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). “Stan: A probabilistic programming language.” *Journal of Statistical Software*, 76(1). 241
- Carvalho, L. M., Villela, D. A. M., Coelho, F. C., and Bastos, L. S. (2022). “Supplementary Material of “Bayesian Inference for the Weights in Logarithmic Pooling”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/22-BA1311SUPP>. 225, 226, 227, 230, 231, 241, 243, 244
- Coelho, F. C. and Codeço, C. T. (2009). “Dynamic modeling of vaccinating behavior as a function of individual beliefs.” *PLoS Comput. Biol.*, 5(7): e1000425. MR2538901. doi: <https://doi.org/10.1371/journal.pcbi.1000425>. 223
- DasGupta, A. (2011). “The exponential family and statistical applications.” In *Probability for Statistics and Machine Learning*, 583–612. Springer. MR2807365. doi: <https://doi.org/10.1007/978-1-4419-9634-3>. 226
- Diaconis, P. and Ylvisaker, D. (1979). “Conjugate priors for exponential families.” *The Annals of Statistics*, 269–281. MR0520238. 227
- French, S. (1985). “Group consensus probability distributions: A critical survey in Bayesian statistics.” *Bayesian Statistics*, 2. MR0862490. 223, 224
- Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P. (2019). *Handbook of Mixture Analysis*. CRC Press. MR3889980. 245
- Genest, C. (1984). “A characterization theorem for externally Bayesian groups.” *Annals*

- of Statistics*, 12(3): 1100–1105. MR0751297. doi: <https://doi.org/10.1214/aos/1176346726>. 225
- Genest, C., McConway, K. J., and Schervish, M. J. (1986). “Characterization of externally Bayesian pooling operators.” *The Annals of Statistics*, 487–501. MR0840510. doi: <https://doi.org/10.1214/aos/1176349934>. 223, 225
- Genest, C., Weerahandi, S., and Zidek, J. V. (1984). “Aggregating opinions through logarithmic pooling.” *Theory and Decision*, 17(1): 61–70. MR0752843. doi: <https://doi.org/10.1007/BF00140056>. 223, 224, 225, 226
- Genest, C. and Zidek, J. V. (1986). “Combining probability distributions: A critique and an annotated bibliography.” *Statistical Science*, 114–135. MR0833278. 223, 244
- Guardoni, G. L. (2002). “On irrelevance of alternatives and opinion pooling.” *Brazilian Journal of Probability and Statistics*, 87–98. MR1962496. 223
- Jackson, C. H., Jit, M., Sharples, L. D., and De Angelis, D. (2015). “Calibration of complex models through Bayesian evidence synthesis: a demonstration and tutorial.” *Medical Decision Making*, 35(2): 148–161. 247
- Jaynes, E. T. (1957). “Information theory and statistical mechanics. II.” *Physical Review*, 108: 171–190. MR0096414. 228
- Jombart, T., Frost, S., Nouvellet, P., Campbell, F., and Sudre, B. (2019). *outbreaks: A Collection of Disease Outbreak Data*. R package version 1.6.0. URL <https://github.com/reconhub/outbreaks> 242
- Leutbecher, M. and Palmer, T. N. (2008). “Ensemble forecasting.” *Journal of Computational Physics*, 227(7): 3515–3539. MR2400226. doi: <https://doi.org/10.1016/j.jcp.2007.02.014>. 243
- Li, Z. S., Guo, J., Xiao, N.-C., and Huang, W. (2017). “Multiple priors integration for reliability estimation using the Bayesian melding method.” In *Reliability and Maintainability Symposium (RAMS), 2017 Annual*, 1–6. IEEE. 230
- Lind, N. C. and Nowak, A. S. (1988). “Pooling expert opinions on probability distributions.” *Journal of Engineering Mechanics*, 114(2): 328–341. 223
- Lindley, D. V. (2013). *Understanding Uncertainty*. John Wiley & Sons. MR3236718. doi: <https://doi.org/10.1002/9781118650158.indsp2>. 246
- McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2019). “Multivariate Bayesian predictive synthesis in macroeconomic forecasting.” *Journal of the American Statistical Association*. arXiv:1711.01667. Published online: Oct 9 2019. MR4143452. doi: <https://doi.org/10.1080/01621459.2019.1660171>. 246
- McAlinn, K., Aastveit, K. A., and West, M. (2018). “Bayesian predictive synthesis – discussion of: Using stacking to average Bayesian predictive distributions, by Y. Yao et al.” *Bayesian Analysis*, 13: 971–973. MR3853125. doi: <https://doi.org/10.1214/17-BA1091>. 246
- McAlinn, K. and West, M. (2019). “Dynamic Bayesian predictive synthesis in time series

- forecasting.” *Journal of Econometrics*, 210: 155–169. arXiv:1601.07463. MR3944768. doi: <https://doi.org/10.1016/j.jeconom.2018.11.010>. 246
- Murray, J. D. (2002). *Mathematical Biology I. An Introduction*, volume 17 of *Interdisciplinary Applied Mathematics*. New York: Springer, 3rd edn. MR1908418. 243, 244
- Myung, I. J., Ramamoorti, S., and Bailey Jr, A. D. (1996). “Maximum entropy aggregation of expert predictions.” *Management Science*, 42(10): 1420–1436. 224
- Neal, R. M. (2003). “Slice sampling.” *The Annals of Statistics*, 31(3): 705–767. MR1994729. doi: <https://doi.org/10.1214/aos/1056562461>. 226
- Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). “A note on the power prior.” *Statistics in Medicine*, 28(28): 3562–3566. MR2744383. doi: <https://doi.org/10.1002/sim.3722>. 228, 241
- Pennock, D. M. and Wellman, M. P. (1997). “Representing aggregate belief through the competitive equilibrium of a securities market.” In Geiger, D. and Shenoy, P. P. (eds.), *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 392–400. Morgan Kaufmann Publishers Inc. 223
- Poole, D. and Raftery, A. E. (2000). “Inference for deterministic simulation models: the Bayesian melding approach.” *Journal of the American Statistical Association*, 95(452): 1244–1255. MR1804247. doi: <https://doi.org/10.2307/2669764>. 223, 224, 225, 228, 230, 238, 239, 240, 241, 243, 246
- Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2007). “Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.” In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics*, 1–45. Oxford University Press. MR2433201. 233
- Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Science & Business Media. MR2723361. 226
- Rufo, M., Martin, J., Pérez, C., et al. (2012a). “Log-linear pool to combine prior distributions: a suggestion for a calibration-based approach.” *Bayesian Analysis*, 7(2): 411–438. MR2934957. doi: <https://doi.org/10.1214/12-BA714>. 224, 228, 229, 233
- Rufo, M. J., Pérez, C. J., Martín, J., et al. (2012b). “A Bayesian approach to aggregate experts’ initial information.” *Electronic Journal of Statistics*, 6: 2362–2382. MR3020268. doi: <https://doi.org/10.1214/12-EJS752>. 224, 228, 232, 233, 234
- Saumard, A. and Wellner, J. A. (2014). “Log-concavity and strong log-concavity: a review.” *Statistics Surveys*, 8: 45. MR3290441. doi: <https://doi.org/10.1214/14-SS107>. 226
- Savchuk, V. P. and Martz, H. F. (1994). “Bayes reliability estimation using multiple sources of prior information: binomial sampling.” *IEEE Transactions on Reliability*, 43(1): 138–144. 223, 228, 232, 233, 234

- Smith, M. R. (2017). “Ternary: An R Package for Creating Ternary Plots.” *Zenodo*. 239
- Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). “Validating Bayesian inference algorithms with simulation-based calibration.” *arXiv preprint arXiv:1804.06788*. 235
- West, M. (1984). “Bayesian aggregation.” *Journal of the Royal Statistical Society. Series A (General)*, 600–607. MR0782071. doi: <https://doi.org/10.2307/2981847>. 223, 244
- Yao, Y., Vehtari, A., Simpson, D., Gelman, A., et al. (2018). “Using stacking to average Bayesian predictive distributions (with discussion).” *Bayesian Analysis*, 13(3): 917–1003. MR3853125. doi: <https://doi.org/10.1214/17-BA1091>. 246
- Zhong, M., Goddard, N., and Sutton, C. (2015). “Latent Bayesian melding for integrating individual and population models.” In *Advances in Neural Information Processing Systems*, 3618–3626. 230

Acknowledgments

The authors would like to thank Professors Adrian Raftery, Christian Genest, Kevin Mcconway and Mike West, as well as Drs. David Poole, Eduardo Mendes and Felipe Figueiredo for helpful discussions. We would also like to thank the two anonymous referees, whose suggestions greatly improved the paper. DAMV and LSB were supported in part by CAPES under CAPES/Cofecub project (N. 833/15). FCC is grateful to Fundação Getúlio Vargas for funding during this project.