On the choice of the weights for the logarithmic pooling of probability distributions

Luiz Max F. de Carvalho[a,b,c], Daniel A. M. Villela[a], Flavio Coelho[c] & Leonardo S. Bastos[a]
a – Program for Scientific Computing (PROCC), Oswaldo Cruz Foundation.
b – Institute of Evolutionary Biology, University of Edinburgh.
c – School of Applied Mathematics, Getulio Vargas Foundation (FGV).

November 20, 2018

Abstract

Combining different prior 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. The resulting pooled distribution however depends heavily on set of weights given to each opinion/prior. In this paper we exploreapproaches to assigning weights to opinions. Two methods are stated in terms of optimisation problems and a third one uses a hierarchical prior that accounts for uncertainty on the weights. We explore several examples of interest, such as proportion and rate estimation and combining Normal distributions. Our findings...

Key-words: logarithmic pooling; expert opinion; hierarchical modelling; Bayesian melding.

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). 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). 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 **Background**). In a practical setting, logarithmic pooling finds use in a 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 $\theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ when they can be represented as (proper) probability distributions. Combining these opinions using logarithmic pooling requires assigning weights to each of the experts. These weights represent the reliability of each opinion (Genest et al., 1984). This requirement naturally leads to the question of how to choose the weights in a meaningful fashion, 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 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).

These approaches, while moving away from the problem of arbitrarily assigning the weights, arrive at single point solutions, similar to point estimates in Statistical theory. Albeit acknowledging that these approaches have merit, we argue that in many settings, where one has substantial prior information on the relative reliabilities of the information sources (experts), it would be desirable to incorporate this information 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 deriving the weights for logarithmic pooling and propose a hierarchical prior approach to learning about the weights. In section 2 we provide some background and notation on logarithmic pooling. In section 3 we present different approaches to choose the weights, two methods based on optimality criteria and one based on hierarchical modelling.

One potential application of logarithmic pooling is in meta analyses. Distributions of a quantity of interest estimated in several studies can be combined in a single consensus distribution in a principled way using logarithmic pooling,. In addition, one can use features of the studies such as the sample sizes to construct the weights. In epidemiology, estimation of disease prevalence and the effect of exposure variables are amongst the most important application of meta-analyses. In this paper, section 4.1, we explore a meta-analysis application where the interest is to pool results from six studies on HIV prevalence in men who have sex with men (MSM) populations in Brazil (Malta et al., 2010).

Another important application of logarithmic pooling is within the Bayesian melding method of Poole and Raftery (2000). The Bayesian melding method concerns drawing inference about a deterministic model by combining a natural prior on quantities of interest (inputs and/or outputs) with the prior *induced* by the model through logarithmic pooling. 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 acknowledgement of uncertainty is crucial. In section4.3 we revisit the non-aged-structured population deterministic model (PDM) for bowhead whales explored in Poole and Raftery (2000) and extend their approach to accommodate uncertainty about the weighting.

2 Background

In what follows, 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 $\mathbf{F}_\theta := \{f_0(\theta), f_1(\theta), f_2(\theta), \dots, f_K(\theta)\}$ be a set of prior 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 $\alpha_i > 0\ \forall i$ and $\sum_{i=0}^K \alpha_i = 1$. Then the log-pooled prior is

$$\mathcal{LP}(\mathbf{F}_\theta, \boldsymbol{\alpha}) := \pi(\theta) = t(\boldsymbol{\alpha}) \prod_{i=0}^{K} f_i(\theta)^{\alpha_i}, \tag{1}$$

where $t(\boldsymbol{\alpha}) = \int_{\boldsymbol{\Theta}} \prod_{i=0}^K f_i(\theta)^{\alpha_i} d\theta$.

Logarithmic pooling will only yield proper probability distributions if it is possible to normalise the expression in (1). While Poole and Raftery (2000) provide a proof for the case of two densities (see Theorem 1 therein), Genest et al. (1986) (pg.489) prove the result for a finite number of densities.

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

A simple proof using Hölder's inequality is given in the Appendix of this paper. This result ensures any (finite) number of proper distributions can be combined using the logarithmic pooling operator to yield a normalisable (proper) density.

Another interesting property of the logarithmic pooling operator is the fact that combining log-concave densities will produce a log-concave pooled distribution.

Remark 1. ***Log-concavity.*** *Let $\mathbf{F}_\theta$ be a set of log-concave distributions, i.e., each f_i can be represented as $f_i(\theta) \propto e^{\psi_i(\theta)}$, where $\psi_i(\cdot)$ is a concave function. Then $\pi(\theta)$ is also log-concave.*

Log-concavity of the pooled prior may be important to consider in order to guarantee unimodality and certain conditions on tail behaviour.

2.1 Exponential family

Suppose we are interested in a random variable, Y, from a exponential family with parameter θ and probability density function given by

$$f(y|\theta) = h(y)e^{\theta y - s(\theta)}. \tag{2}$$

Let $\mathbf{F}_y$ be a set of distributions on y of the form in (2), $f_i(y|\theta_i)$, $i = 0, 1, \dots, K$. The combined (log-pooled) distribution also belongs to the exponential family:

$$\pi(y|\boldsymbol{\alpha}) = t(\boldsymbol{\alpha})h^*(y)e^{\theta^* y - s^*(\boldsymbol{\theta})}. \tag{3}$$

where $\boldsymbol{\theta} := \{\theta_0, \theta_1, \dots, \theta_K\}$, $h^*(y) = \prod_{i=0}^K h_i(y)^{\alpha_i}$, $\theta^* = \sum_{i=0}^K \alpha_i\theta_i$ and $s^*(\boldsymbol{\theta}) = \sum_{i=0}^K \alpha_i s_i(\theta_i)$.

The entropy function of the log-pooled distribution is

$$H_\pi(Y;\boldsymbol{\alpha}) := -\mathbb{E}_\pi\left[-\log \pi(Y|\boldsymbol{\alpha})\right] = -\log t(\boldsymbol{\alpha}) + s^*(\boldsymbol{\theta}) - \mathbb{E}_\pi[\log h^*(Y)] - \theta^*\mathbb{E}_\pi[Y]\,, \tag{4}$$

where $\mathbb{E}_\pi\left[g(Y)\right]$ is the expectation of $g(Y)$ given that Y has a probability density function $\pi(y)$, i.e. $\mathbb{E}_\pi\left[g(Y)\right] = \int_{-\infty}^{\infty} g(y)\pi(y)dy$.

The Kullback-Leibler divergence between the pooled distribution (3) and each distribution in $\mathbf{F}_y$ can be written as:

$$KL(\pi||f_i) = -H_\pi(Y;\boldsymbol{\alpha}) - \mathbb{E}_\pi[\log h_i(Y)] - \theta_i\mathbb{E}_\pi[Y] + s_i(\theta_i)\,. \tag{5}$$

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.

2.2 Conjugate priors to the exponential family

A conjugate prior family for $f(y|\theta)$, given in (2), is the following

$$g(\theta|a,b) = K(a,b)e^{\theta a - bs(\theta)}\,, \tag{6}$$

where $K(a,b)$ is a normalising constant. Here, we employ the definition and notation for exponential family and its conjugate prior family from Robert (2001, chapter 3). Similar to the above, let $\mathbf{G}_\theta$ be a set of conjugate prior distributions representing the opinions of $K+1$ experts, and $g_i(\theta) = g(\theta|a_i,b_i)$ from equation (6).

The log-pooled prior is also a conjugate prior for $f(y|\theta)$ with hyperparameters given by an weighted mean of the experts hyperparameters, i.e., $\pi(\theta|\boldsymbol{\alpha}) = g(\theta|a^*,b^*)$, 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 (6) is given by

$$H_\pi(\theta;\boldsymbol{\alpha}) = -\log(K(a^*,b^*)) - a^*\mathbb{E}_\pi[\theta] + b^*\mathbb{E}_\pi[s(\theta)]\,. \tag{7}$$

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

$$KL(\pi||g_i) = -H_\pi(\theta;\boldsymbol{\alpha}) - \log(K(a_i,b_i)) - a_i\mathbb{E}_\pi[\theta] + b_i\mathbb{E}_\pi[s(\theta)]\,. \tag{8}$$

We now move on to study three approaches to assign weights. The first two approaches are based on optimality criteria and a third method is based on assigning a Dirichlet hyperprior on the weights.

3 Assigning the weights in logarithmic pooling

3.1 Choosing weights based on optimality criteria

3.1.1 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 uninformative. Such approach would ensure that, given the constraints imposed by $\mathbf{F}_\theta$, 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. Formally, we want to find $\hat{\boldsymbol{\alpha}}$ such that

$$\hat{\boldsymbol{\alpha}} := \arg\max H_\pi(\theta;\boldsymbol{\alpha}). \tag{9}$$

This approach, however, does not result in a convex optimisation problem, therefore one is not guaranteed to find a unique solution. See Proposition 2, below, for intuition as to why.

3.1.2 Minimising Kullback-Leibler divergence

One could also wish to choose the pooling weights so as to minimise the total Kullback-Leibler divergence between the pooled distribution, π, each proposed distribution in $\mathbf{F}_\theta$. Let $d_i = \mathrm{KL}(\pi||f_i)$ and let $L(\boldsymbol{\alpha})$ be a loss function such that

$$L(\boldsymbol{\alpha}) = \sum_{i=0}^{K} d_i \tag{10}$$

$$= -(K+1)H_\pi(\theta;\boldsymbol{\alpha}) - \sum_{i=0}^{K} \mathbb{E}_\pi \left[\log f_i(\theta)\right] \tag{11}$$

$$\hat{\boldsymbol{\alpha}} := \arg\min L(\boldsymbol{\alpha}) \tag{12}$$

[DO WE NEED UNIQUENESS?]

If one want to look at the individual weights, then uniqueness would be an important property.

Remark 2. ***Uniqueness*** *(Rufo et al., 2012a). The distribution obtained following (12) is unique, i.e., there is only one aggregated prior $\pi(\theta)$ that minimizes $L(\boldsymbol{\alpha})$.*

One can get some intuition into the proof by noting that minimising (11) is equivalent to maximising $\ln t(\boldsymbol{\alpha}) = \ln \int_{\boldsymbol{\Theta}} \prod_{i=0}^{K} f_i(\theta)^{\alpha_i} d\theta$. Rufo et al. (2012a) show that $t(\boldsymbol{\alpha})$ is concave, therefore the problem in (12) has a unique solution. By contrast, the problem in (9) requires to minimise $\ln t(\boldsymbol{\alpha})$ hence lacking a sufficient condition for the existence of a unique solution.

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 doesn't scale well with increasing the number of distributions being combined and; (b) it fails to account for any (posterior) dependency between model parameters and the weights. In this section we propose assigning 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.

$$\pi_D(\boldsymbol{\alpha}) = \frac{1}{\mathcal{B}(\boldsymbol{x})} \prod_{i=0}^{K} \alpha_i^{x_i - 1} \tag{13}$$

where $\boldsymbol{x} = \{x_0, x_1, \ldots, x_K\}$ is the vector of hyperparameters for the Dirichlet prior and $\mathcal{B}(X)$ 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).

$$\pi_A(\boldsymbol{\alpha}) = \frac{1}{|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}} \frac{1}{\prod_{i=0}^{K}\alpha_i} e^{\left(\log\left(\frac{\boldsymbol{\alpha}_{-K}}{\alpha_K}\right) - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\log\left(\frac{\boldsymbol{\alpha}_{-K}}{\alpha_K}\right) - \boldsymbol{\mu}\right)} \tag{14}$$

where $\boldsymbol{\alpha}_{-K}$ represents the vector $\boldsymbol{\alpha}$ 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 (13) and the logistic-normal (14) distributions, i.e.

$$\mu_i = \psi(x_i) - \psi(x_K), \quad i = 0, 1, \ldots, K-1, \tag{15}$$

$$\Sigma_{ii} = \psi'(x_i) + \psi'(x_K), \quad i = 0, 1, \ldots, K-1, \tag{16}$$

$$\Sigma_{ij} = \psi'(x_K) \tag{17}$$

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

4 Applications

4.1 HIV prevalence among MSM populations in Brazil

In epidemiology, systematic review and meta analysis are popular tools for merging and contrasting results across multiple studies (Rothman et al., 2008, Chapter 33). For instance, the logarithmic polling could be used to combine probability distributions of a particular outcome estimated from several studies.

We illustrate the different approaches to assign weights in the logarithmic polling in the systematic review and meta analysis conducted by Malta et al. (2010). They analysed studies published from 1999 to 2009 assessing the HIV prevalence among men who have sex with another men (MSM) in Brazil. The authors have found six studies that estimated HIV prevalence in MSM population in Brazil.

Assuming a uniform prior for the HIV prevalence among MSM, denoted by θ, and a binomial model for each study, i.e. $Y_i \sim Binom(n_i, \theta)$. The posterior distribution for the HIV prevalence conditional on each study is then a Beta distribution with parameters $a_i = y_i + 1$ and $b_i = n_i - y_i + 1$, for $i = 0, 1, 2, 3, 4, 5$. The probability density of the HIV prevalence for each study is given by

$$f_i(\theta; a_i, b_i) = \frac{1}{B(a_i, b_i)} \theta^{a_i - 1}(1-\theta)^{b_i - 1},$$

where $B(a, b) := \int_0^1 x^{a-1}(1-x)^{b-1}dx$ is the Beta function.

Table 1 presents for each study the sample size, the total of HIV positive observed, and the estimated prevalence given by the expectation of the Beta distribution with 95% credible interval. Note that the estimated prevalences among MSM are very high when compared with the HIV prevalence in the general population, 0.6% (Malta et al., 2010).

Table 1: Data extracted from the systematic review and meta analysis conducted by Malta et al. (2010) assessing the HIV prevalence among MSM in Brazil. n is the sample size, y is the total of HIV-positive participants.

Study	Reference	n	y	Estimated prevalence (95% CI)
0	Tun et al. (2008)	658	44	0.0682 (0.0502, 0.0886)
1	Barcellos et al. (2003)	461	111	0.2419 (0.2040, 0.2819)
2	Carneiro et al. (2003)	621	61	0.0995 (0.0773, 0.1242)
3	Sutmoller et al. (2002)	1165	281	0.2416 (0.2175, 0.2666)
4	Brazilian Ministry of Health (2000)	642	57	0.0901 (0.0692, 0.1133)
5	Harrison et al. (1999)	849	99	0.1175 (0.0968, 0.1400)

The log-pooled distribution for the HIV prevalence is then

$$\begin{aligned} \pi(\theta) &= t(\boldsymbol{\alpha}) \prod_{i=0}^{K} f_i(\theta; a_i, b_i)^{\alpha_i} && (18) \\ &\propto \prod_{i=0}^{K} \left(\theta^{a_i-1}(1-\theta)^{b_i-1}\right)^{\alpha_i} && (19) \\ &\propto \theta^{a^*-1}(1-\theta)^{b^*-1} && (20) \end{aligned}$$

with $a^* = \sum_{i=0}^{K} \alpha_i a_i$ and $b^* = \sum_{i=0}^{K} \alpha_i b_i$. Note that (20) is the kernel of a Beta distribution with parameters a^* and b^*. Hence the entropy is the following

$$H_\pi(\theta) = \log B(a^*, b^*) - (a^*-1)\psi(a^*) - (b^*-1)\psi(b^*) + (a^*+b^*-2)\psi(a^*+b^*). \tag{21}$$

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

$$\begin{aligned} d_i = KL(\pi || f_i) = \ln\left(\frac{\mathcal{B}(a_i, b_i)}{\mathcal{B}(a^*, b^*)}\right) &+ (a^* - a_i)\psi(a^*) + (b^* - b_i)\psi(b^*) \\ &- (a^* - a_i + b^* - b_i)\psi(a^* + b^*). \end{aligned} \tag{22}$$

Table 2 shows the estimates of the HIV prevalence among MSM in Brazil using different methods of dealing with weights. The first method assumes all studies are equally important, i.e. $\alpha_i = \frac{1}{K}, \quad \forall i$. The second method maximizes the entropy given in (21) which in this example corresponds to ($\alpha_3 = 1, \alpha_j = 0 \quad \forall j \neq 3$). This method gave all the weight for the study 3 by Sutmoller et al. (2002) which is the study with the larger sample size. The remaining methods, minimizing the KL divergence and the hierarchical modelling of weights suggest that all studies should have the same weight, however in the hierarchical approach the weights probability distribution is numerically integrated out leading to larger 95% credible intervals.

Table 2: Estimated HIV prevalence (95% credible interval) among men who have sex with another man in Brazil using different methods for assessing the weights.

Method	Estimated HIV prevalence
Equal weights	0.1495 (0.1247, 0.1762)
Maximum entropy	0.2416 (0.2175, 0.2666)
Minimum KL	0.1495 (0.1247, 0.1762)
Hierarchical Dirichlet prior	0.1478 (0.0941, 0.2117)
Hierarchical Aitchison prior	0.1471 (0.0878, 0.2245)

4.2 Combining expert priors on failure probabilities [MELHORAR]

We now turn our attention to combining expert opinions about probabilities and proportions. Here we analyse an example proposed by Savchuk and Martz (1994) (also discussed in Rufo et al. (2012b)) in which four experts are required supply prior information about the survival probability θ of a certain unit, for which there have been $y = 9$ successes out of $n = 10$ trials. 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 variance 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 $\mathbf{b} = \{b_0 = 0.955, b_1 = 0.860, b_2 = 0.924, b_3 = 0.848\}$. Note that this example is quite similar to the previous one in the sense that the aim is to combine beta distributions. However, we are able to estimate the posterior distribution for the survival probabilities and also, in the hierarchical modelling approach, the posterior distribution for the weights.

Table 3 lists the weights proposed by each method. Figure 1 shows the prior and posterior distributions in each of the methods and also the case in which we assign an equal weight $(1/K)$ to each opinion. It is interesting to note that maximum entropy suggests to discard all opinions but one, which effectively leads to the maximum entropy. Since $t(\boldsymbol{\alpha})$ is concave, we expect to find the maximum entropy given by the boundary conditions, which may lead to border points in the simplex. Minimising Kullback-Leibler divergence between each prior and the pooled prior leads to finding a unique solution but in this case also suggests to discard two of the opinions. [NEEDS UPDATE!] By contrast, using a hierarchical Dirichlet prior for the weights gives rather different results from the first two methods in proposing almost equal weights to each of the opinions.

Table 3: Weights obtained using the three methods for the proportion estimation problem. [1] – Kullback-Leibler [2] – Posterior mean for $\boldsymbol{\alpha}$.

Method	α_0	α_1	α_2	α_3
Maximum entropy	0.00	1.00	0.00	0.00
Minimum KL[1] divergence	0.04	0.96	0.00	0.00
Hierarchical prior[2]	0.26	0.24	0.26	0.23

To complete the analysis, we place a diffuse $Dirichlet(\boldsymbol{\alpha}|\boldsymbol{X})$ prior on $\boldsymbol{\alpha}$ with $X_i = 1/4\,\forall i$. Finally, we propose to compare the prior distributions representing the experts' opinions as well as the combined distributions obtained by the different approaches using the integrated (marginal) likelihood (Raftery et al. (2007), eq. 9), $l(y) = \int_0^1 f(y|\theta)\pi(\theta)d\theta$. The marginal likelihood for the $i-th$ expert and J observations of the form $\{y_j, n_j\}$ is:

$$
\begin{aligned}
l_i(y_j) &= \int_0^1 \mathcal{L}(\theta|y_j, n_j)\pi_i(\theta)d\theta \\
&= \prod_{j=1}^{J} \frac{\Gamma(n_j - 1)}{\Gamma(n_j - y_j + 1)\Gamma(y_j + 1)} \frac{\Gamma(a_i + b_i)}{\Gamma(a_i + b_i + n_j)} \frac{\Gamma(a_i + y_j)}{\Gamma(a_i)} \frac{\Gamma(b_i + n_j - y_j)}{\Gamma(b_i)} \qquad (23)
\end{aligned}
$$

One can get insight into these results by looking at the integrated likelihoods in Table 4 and the densities in Figure 1, we note that all three methods lead to similar pooled distributions. Note that the only distribution with a substantially different $l(y)$ is that of Expert 3, who gave a rather divergent mean for the survival probability ($m_3 = 0.70$).

Table 4: Integrated likelihoods ($l(y)$) for the priors of each expert as well as the combined priors. [1] Calculated using the posterior mean of $\boldsymbol{\alpha}$

Expert priors		Pooled priors	
Expert 0	0.237	Equal weights	0.254
Expert 1	0.211	Maximum entropy	0.211
Expert 2	0.256	Minimum KL	0.223
Expert 3	0.163	Hierarchical[1]	0.255

In conclusion, if the prior distributions (opinions) are not radically different, all three optmisation-based methods will likely lead to similar combined priors. Although this is the case for the simple univariate example presented, it remains to be seen if this is the case for high-dimensional θ under complex sampling distributions. As the results presented in this paper make clear, future research shall be focused on cases where there is substantial heterogeneity (disagreement) in the available opinions.

Method	Prior	Posterior
Equal weights	0.90 (0.64–1.00)	0.90 (0.73–0.99)
Maximum entropy	0.80 (0.37–1.00)	0.87 (0.66–0.98)
Minimum KL	0.82 (0.42–1.00)	0.87 (0.67–0.99)
Hierarchical Dirichlet prior	0.88 (0.53–1.00)	0.90 (0.71–0.99)
Hierarchical Aitchinson prior	? (?–?)	? (?–?)

4.3 Bayesian melding with varying weights: bowhead whale population growth

In their seminal paper, Poole and Raftery (2000) propose an application of Bayesian melding to the analysis of a deterministic population model for bowhead whales. The model presented in that paper 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:

$$P_{t+1} = P_t - C_t \times \text{MSYR} \times P_t \left(1 - (P_t/P_0)^2\right). \tag{24}$$

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, such that $P_{1993} = P_{1978}(1 + \text{ROI})^{15}$. Consider the model outputs $\phi = \{P_{1993}, \text{ROI}\}$. Bayesian melding seeks to draw inference by first constructing a prior on ϕ of the form

$$\tilde{q}_\Phi(\phi) \propto q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} \tag{25}$$

where $q_1^*()$ is the **induced** prior on the outputs and $q_2()$ is the prior on ϕ without considering the deterministic model. The prior (25) can then be inverted to obtain a *coherised* prior on θ, $\tilde{q}_\Theta(\theta)$.

Standard Bayesian inference may then follow, leading to the posterior

$$\pi_\Theta(\theta) \propto \tilde{q}_\Theta(\theta) L_1(\theta) L_2(M(\theta)) \tag{26}$$

For details on priors and likelihoods, we refer the reader to the Appendix and Poole and Raftery (2000).

In original paper, (Poole and Raftery, 2000, sec. 3.4) propose a sampling-importance-resampling (SpIR) algorithm to sample from the posterior in (26), which we extend here to in order to accommodate varying weights.

1. Draw k values from $q_1(\theta)$, constructing $\boldsymbol{\theta}_k = (\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(k)})$;
2. Similarly, construct $\boldsymbol{\alpha}_k$ from $\pi(\boldsymbol{\alpha})$;
3. For each $\theta^{(i)} \in \boldsymbol{\theta}_k$ run the model to compute $\psi^{(i)} = M(\theta^{(i)})$, constructing $\boldsymbol{\phi}_k$;
4. Obtain a density estimate of $q_1^\star(\phi)$ from $\boldsymbol{\phi}_k$;

5. Form the importance weights

$$w_i = t(\boldsymbol{\alpha}^{(i)}) \left(\frac{q_2(M(\theta^{(i)}))}{q_1^{\star}(M(\theta^{(i)}))} \right)^{1-\boldsymbol{\alpha}^{(i)}} L_1(\theta^{(i)}) L_2(M(\theta^{(i)})), \tag{27}$$

where $t(\boldsymbol{\alpha}^{(i)})$ is $\int_{\Phi} q_1^*(\phi)^{\alpha^{(i)}} q_2(\phi)^{1-\alpha^{(i)}} d\phi$, computed using standard numerical quadrature methods;

6. (Re)Sample l values from $\boldsymbol{\theta}_k$ according to the weights $\boldsymbol{w}_k$.

5 Discussion and conclusions

6 Acknowledgements

The authors would like to thank Professor Adrian Raftery (University of Washington) and David Poole for helpful suggestions. DAMV and LSB were supported in part by Capes under Capes/Cofecub project (N. 833/15). FCC is grateful to Fundação Getulio Vargas for funding during this project.

References

Abbas, A. E. (2009). A Kullback-Leibler view of linear and log-linear pools. *Decision Analysis*, 6(1):25–37.

Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272.

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.

Barcellos, N. T., Fuchs, S. C., and Fuchs, F. D. (2003). Prevalence of and risk factors for HIV infection in individuals testing for HIV at counseling centers in Brazil. *Sexually transmitted diseases*, 30(2):166–173.

Brazilian Ministry of Health (2000). Bela vista & horizon: Behavioral and serological studies among men who have sex with men. Technical report, Brasília: Brazilian Ministry of Health. In Portuguese.

Carneiro, M., Cardoso, F. A., Greco, M., Oliveira, E., Andrade, J., Greco, D. B., and Antunes, C. M. d. F. (2003). Determinants of human immunodeficiency virus (HIV) prevalence in homosexual and bisexual men screened for admission to a cohort study of HIV negatives in Belo Horizonte, Brazil: Project Horizonte. *Memórias do Instituto Oswaldo Cruz*, 98(3):325–329.

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.

Genest, C., McConway, K. J., and Schervish, M. J. (1986). Characterization of externally bayesian pooling operators. *The Annals of Statistics*, pages 487–501.

Genest, C., Weerahandi, S., and Zidek, J. V. (1984). Aggregating opinions through logarithmic pooling. *Theory and Decision*, 17(1):61–70.

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, pages 114–135.

Harrison, L. H., do Lago, R. F., Friedman, R. K., Rodrigues, J., Santos, E. M., de Melo, M. F., Moulton, L. H., Schechter, M., Group, P. P. O. S., et al. (1999). Incident hiv infection in a high-risk, homosexual, male cohort in rio de janeiro, brazil. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 21(5):408–412.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. II. *Physical Review*, 108:171–190.

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*, pages 1–6. IEEE.

Lind, N. C. and Nowak, A. S. (1988). Pooling expert opinions on probability distributions. *Journal of engineering mechanics*, 114(2):328–341.

Malta, M., Magnanini, M. M., Mello, M. B., Pascom, A. R. P., Linhares, Y., and Bastos, F. I. (2010). HIV prevalence among female sex workers, drug users and men who have sex with men in Brazil: a systematic review and meta-analysis. *BMC Public Health*, 10(1):1.

Myung, I. J., Ramamoorti, S., and Bailey Jr, A. D. (1996). Maximum entropy aggregation of expert predictions. *Management Science*, 42(10):1420–1436.

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.

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., editors, *Bayesian Statistics*, pages 1–45. Oxford University Press.

Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (Springer Texts in Statistics) by.* Springer-Verlag New York.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology.* Lippincott Williams & Wilkins, 3rd. edition.

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.

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.

Savchuk, V. P. and Martz, H. F. (1994). Bayes reliability estimation using multiple sources of prior information: binomial sampling. *Reliability, IEEE Transactions on*, 43(1):138–144.

Sutmoller, F., Penna, T. L., de Souza, C. T. V., Lambert, J., Group, O. C. F. S. P., et al. (2002). Human immunodeficiency virus incidence and risk behavior in the 'Projeto Rio': Results of the first 5 years of the Rio de Janeiro open cohort of homosexual and bisexual men, 1994–1998. *International journal of infectious diseases*, 6(4):259–265.

Tun, W., de Mello, M., Pinho, A., Chinaglia, M., and Diaz, J. (2008). Sexual risk behaviours and HIV seroprevalence among male sex workers who have sex with men and non-sex workers in Campinas, Brazil. *Sexually Transmitted Infections*, 84(6):455–457.

West, M. (1984). Bayesian aggregation. *Journal of the Royal Statistical Society. Series A (General)*, pages 600–607.

Yeh, C.-C. (2011). Hölder's inequality and related inequalities in probability. *International Journal of Artificial Life Research (IJALR)*, 2(1):54–61.

Zhong, M., Goddard, N., and Sutton, C. (2015). Latent bayesian melding for integrating individual and population models. In *Advances in Neural Information Processing Systems*, pages 3618–3626.

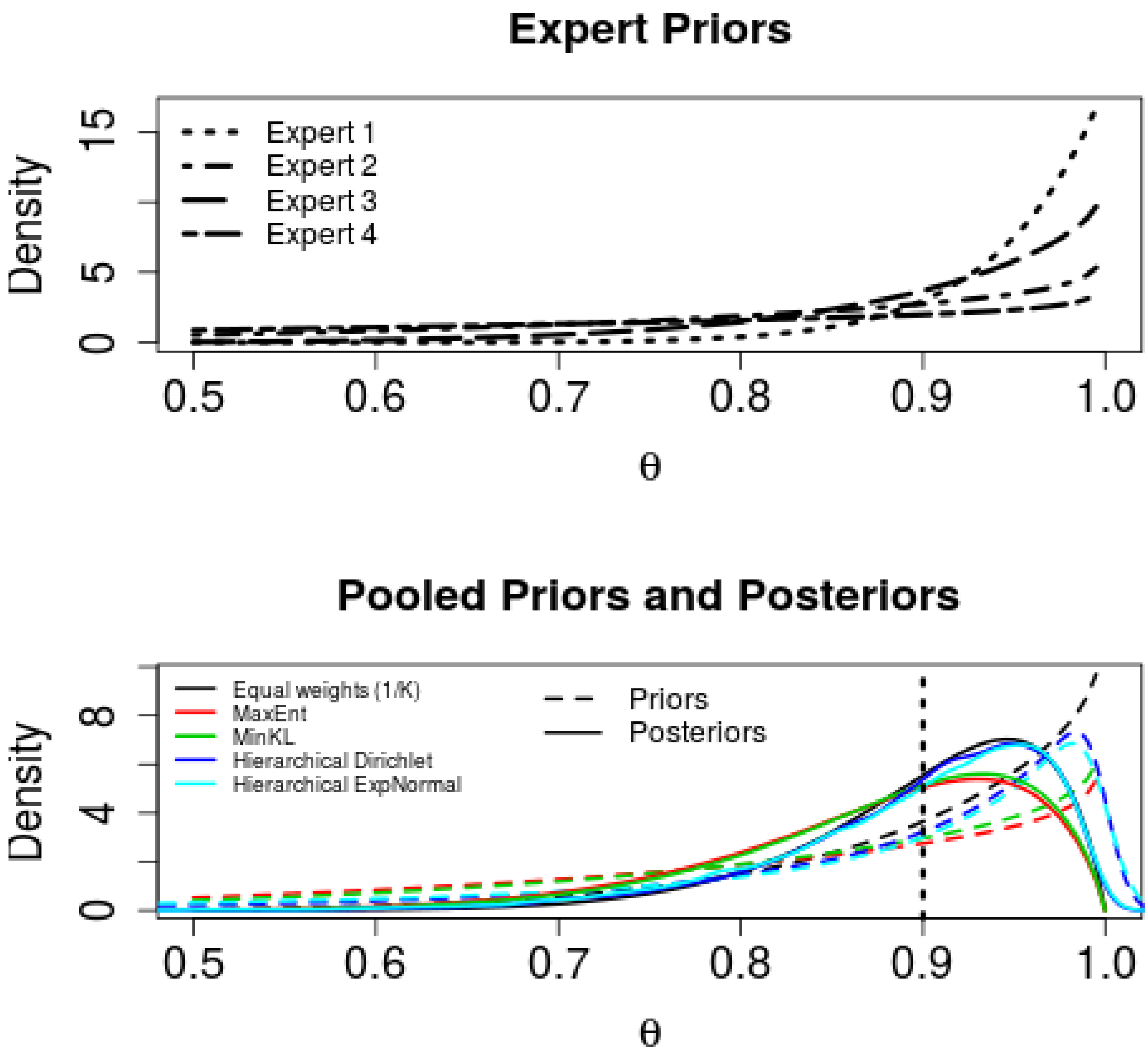

Figure 1: **Prior and posterior densities for θ.** Top panel shows the distributions elicited by each expert (data from Savchuk and Martz (1994)) and the bottom panel shows the pooled priors and posteriors obtained using each of the three methods discussed in this paper. The dashed vertical line marks the maximum likelihood estimate of θ, $\hat{\theta} = 9/10$. [NEEDS CHANGING]

7 Appendix

7.1 Proofs

Here we provide a simple proof of theorem 1 using Hölder's inequality.

Proof. We begin by noting that $\pi(\theta)$ can be re-written as:

$$\pi(\theta) \propto f_0(\theta) \prod_{j=1}^{K} \left(\frac{f_j(\theta)}{f_0(\theta)} \right)^{\alpha_j}. \tag{28}$$

Let $X_j = \frac{f_j(\theta)}{f_0(\theta)}, j = 1, 2, \ldots, K$. Then integrating the expression in (28) is equivalent to finding

$$\mathbb{E}_0 \left[\prod_{j=1}^{K} X_j^{\alpha_j} \right] \leq \prod_{j=1}^{K} \mathbb{E}_0[X_j]^{\alpha_j}, \tag{29}$$

where $\mathbb{E}_0[\cdot]$ is the expectation w.r.t f_0 and (29) follows from Hölder's inequality for expectations (Yeh, 2011). Since $\forall j$ we have $\mathbb{E}_0[X_j]^{\alpha_j} = \left(\int_{\boldsymbol{\Theta}} f_0(\theta)\frac{f_j(\theta)}{f_0(\theta)}\right)^{\alpha_j} d\theta = 1^{\alpha_j} = 1$, Theorem 1 is proven. □

It is straightforward to show that remark 1 holds:

Proof.

$$\pi(\theta) \propto \prod_{i=0}^{K}[\exp(\psi_i(\theta))]^{\alpha_i} \tag{30}$$

$$\propto \exp(\psi^*(\theta)), \tag{31}$$

where $\psi^*(\theta) = \sum_{i=0}^{K} \alpha_i\psi_i(\theta)$ is a concave function due to being a linear combination of concave functions. □

To see that equation 7 holds:

$$\begin{aligned}
H_\pi(\theta) &= \mathbb{E}[-\log(\pi(\theta)] \\
&= -\int \log(\pi(\theta)\pi(\theta)d\theta \\
&= -\int (\log(K(a^*, b^*) + \theta a^* - \psi(\theta)b^*)\pi(\theta)d\theta \\
&= -\log(K(a^*, b^*) + a^*\mathbb{E}[\theta] - b^*\mathbb{E}[\psi(\theta)]
\end{aligned}$$

Likewise for equation 8, we have

$$\begin{aligned}
KL(f_i||\pi) &= \mathbb{E}_\pi[\log(f_i(\theta) - \log(\pi(\theta)] \\
&= \int [\log(K(a_i, b_i)e^{\theta a_i - b_i\psi(\theta)}) - \log(K(a^*, b^*)e^{\theta a^* - b^*\psi(\theta)})]\pi(\theta)d\theta \\
&= \int [\log(K(a_i, b_i)) - \log(K(a^*, b^*)) + (a_i - a^*)\theta - (b_i - b^*)\psi(\theta)\pi(\theta)d\theta \\
&= \log(K(a_i, b_i)) - \log(K(a^*, b^*)) + (a_i - a^*)\mathbb{E}[\theta] - (b_i - b^*)\mathbb{E}[\psi(\theta)]
\end{aligned}$$

Pooling of common distributions

Gamma distributions

Suppose we are interested in a certain count $Y \sim Poisson(\lambda)$ and $K+1$ experts are called upon to elicit prior distributions for λ. A convenient parametric choice for $\mathbf{F}(\lambda)$ is the Gamma family of distributions, for which densities are of the form

$$f_i(\lambda; a_i, b_i) = \frac{b_i^{a_i}}{\Gamma(a_i)}\lambda^{a_i-1}e^{-b_i\lambda}$$

The log-pooled prior $\pi(\lambda)$ is then

$$\pi(\lambda) = t(\boldsymbol{\alpha})\prod_{i=0}^{K} f_i(\lambda; a_i, b_i)^{\alpha_i} \tag{32}$$

$$\propto \prod_{i=0}^{K}\left(\lambda^{a_i-1}e^{-b_i\lambda}\right)^{\alpha_i} \tag{33}$$

$$\propto \lambda^{a^*-1}e^{-b^*\lambda} \tag{34}$$

where $a^* = \sum_{i=0}^{K}\alpha_i a_i$ and $b^* = \sum_{i=0}^{K}\alpha_i b_i$. Noticing (34) is the kernel of a gamma distribution with parameters a^* and b^*, $H_\pi(\lambda)$ becomes

$$H_\pi(\lambda) = a^* - \ln b^* + \ln\Gamma(a^*) + (1 - a^*)\psi(a^*) \tag{35}$$

where $\psi(\cdot)$ is the digamma function. The Kullback-Liebler divergence between each density and the pooled density is:

$$d_i = \mathrm{KL}(f_i||\pi) = (a_i - a^*)\psi(a_i) - \ln\Gamma(a_i) + \ln\Gamma(a^*) + a^*(\ln\frac{b_i}{b^*}) + a_i\frac{b^* - b_i}{b_i} \tag{36}$$

Gaussian distributions

Now suppose one is interested in combinining prior distributions on a quantity $\mu \in \mathbb{R}$. Suppose further that the expert priors are densities in the normal family, i.e.,

$$f_i(\mu; m_i, s_i) = \frac{1}{s_i\sqrt{2\pi}} \exp\left(\frac{-(\mu - m_i)^2}{2s_i^2}\right)$$

Thus

$$\begin{aligned}
\pi(\mu) &= t(\boldsymbol{\alpha}) \prod_{i=0}^{K} f_i(\mu; m_i, s_i)^{\alpha_i} && (37)\\
&\propto \prod_{i=0}^{K} \left[\exp\left(\frac{-(\mu - m_i)^2}{2s_i^2}\right)\right]^{\alpha_i} && (38)\\
&\propto \exp\left[-\frac{1}{2}\left\{\mu \sum_{i=0}^{K} \frac{\alpha_i}{s_i^2} - 2\mu \sum_{i=0}^{K} \frac{\alpha_i m_i}{s_i^2} - \sum_{i=0}^{K} \frac{\alpha_i m_i^2}{s_i^2}\right\}\right] && (39)
\end{aligned}$$

which yields a normal distribution with parameters and $m^* = \frac{\sum_{i=0}^{K} w_i m_i}{\sum_{i=0}^{K} w_i}$ and $v^* = [\sum_{i=0}^{K} w_i]^{-1}$, where $w_i = \alpha_i / s_i^2$. The entropy function is then:

$$H_\pi(\mu) = \frac{1}{2}\left[\ln(2\pi e) - \ln \sum_{i=0}^{K} w_i\right] \qquad (40)$$

which achieves its maximum when $\alpha_j = 1$ for $s_j = max(s_1, s_2, \ldots, s_K)$.