A Appendix

A.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:

$$\pi(\theta) \propto f_0(\theta) \prod_{j=1}^{K} \left(\frac{f_j(\theta)}{f_0(\theta)} \right)^{\alpha_j}. \tag{38}$$

Let $X_j = \frac{f_j(\theta)}{f_0(\theta)}, j = 1, 2, \ldots, K$. Then, integrating the expression in (38) is equivalent to finding

$$\mathbb{E}_0 \left[\prod_{j=1}^{K} X_j^{\alpha_j} \right] \leq \prod_{j=1}^{K} \mathbb{E}_0[X_j]^{\alpha_j}, \tag{39}$$

where $\mathbb{E}_0[\cdot]$ is the expectation w.r.t f_0 and (39) follows from Hölder's inequality for expectations (Yeh, 2011). Since $\forall j$ we have $\mathbb{E}_0[X_j]^{\alpha_j} = \left(\int_{\boldsymbol{\Theta}} f_0(\theta) \frac{f_j(\theta)}{f_0(\theta)} \, d\theta \right)^{\alpha_j} = 1^{\alpha_j}$, Theorem 1 is proven. □

To establish Theorem 2, we will need the following result from Genest et al. (1984).

Lemma 1. ***Representation of a pooling operator with RPC*** *(Genest et al., 1984, eq. 3.1). The only relative propensity consistent operator can always be represented by*

$$\mathcal{T}\left(\boldsymbol{F}_\theta\right)(\theta) = \boldsymbol{B}\left(\boldsymbol{F}_\theta\right) c(\theta) \prod_{i=0}^{K} \left[f_i(\theta) \right]^{\alpha_i}, \tag{40}$$

with $\boldsymbol{B}\left(\boldsymbol{F}_\theta\right) > 0$, $c(\theta) > 0$ *and* $\alpha_0, \alpha_1, \ldots, \alpha_K \geq 0$.

We refer the reader to Genest et al. (1984) for the proof. In short, Lemma 1 is a uniqueness result; logarithmic pooling is the only operator with RPC (Remark 2) and every operator with RPC can be represented as in (40). Now we can state the proof of Theorem 2.

Proof. First, we will show by direct calculation that logarithmic pooling (LP) leads to a log-concave distribution. Notice that each f_i can be written as $f_i(\theta) \propto e^{\nu_i(\theta)}$, where $\nu_i(\cdot)$ is a concave function. We can then write

$$\begin{aligned} \pi(\theta \mid \boldsymbol{\alpha}) &\propto \prod_{i=0}^{K} [\exp(\nu_i(\theta))]^{\alpha_i}, \\ &\propto \exp(\nu^*(\theta)), \end{aligned}$$

where $\nu^*(\theta) = \sum_{i=0}^{K} \alpha_i \nu_i(\theta)$ is a concave function because it is a linear combination of concave functions.

We will now show that LP is the only operator that guarantees log-concavity when $\boldsymbol{F}_\theta$ is a set of concave distributions. First, recall that LP is the only pooling operator that enjoys RPC (Remark 2). With the goal of obtaining a contradiction, suppose that there exists a pooling operator $\mathcal{T}$ which is log-concave but does not enjoy RPC. From Lemma 1, we know that, under this assumption, $\mathcal{T}$ cannot be written in the form $\boldsymbol{B}(\boldsymbol{F}_\theta) c(\theta) \prod_{i=0}^{K} f_i(\theta)^{\alpha_i}$. But every non-negative log-concave function $g(\theta)$ can be represented as

$$g(\theta) = a \cdot c(\theta) \cdot h(\theta), \tag{41}$$

with $a \geq 0$ and $c(\theta)$ and $h(\theta)$ non-negative and log-concave, but otherwise arbitrary. Under the assumptions on $\boldsymbol{F}_\theta$, we have that $h(\theta) := \prod_{i=0}^{K} f_i(\theta)^{\alpha_i}$ is non-negative and log-concave. The only restriction on $c(\theta)$ is that it be positive, it may very well be (log-)concave. Therefore $\mathcal{T}$ can in fact be represented in the form of Lemma 1, contradicting our initial assumption that $\mathcal{T}$ can at same time be log-concave but not enjoy RPC. □

We now move on to the exponential family results in Section 2.1. To see that equation (10) holds:

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

Likewise for equation (11), we have

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

A.2 Computational details

The analyses presented in this paper necessitated numerical optimisation to find the weights based on optimality criteria and Markov chain Monte Carlo (MCMC) to approximate posterior distributions. All computations were carried out in the R (R Core Team, 2019) statistical computing environment, version 3.6.0. We provide implementations using the Stan (Carpenter et al., 2017) probabilistic programming language and R code for the methods, figures and tables presented in this paper can also be found at [blinded for review, please request code if necessary]

A.2.1 Optimisation procedures

In this section we give more detail on the optimisation procedures used to solve the problems in Sections 3.1.1 and 3.1.2. Since both problems are numerically unstable, we employ an strategy that starts the optimisation routine from $J = 1000$ overdispersed points in the unconstrained space, $\mathbb{R}^K$, obtained through the unit simplex transform (Betancourt, 2012). The procedure then picks the overall lowest/highest optimised value in order to avoid local minima/maxima. We draw the initial values from a normal distribution with mean 0 and variance 100^2 and then employ the `optim()` function to optimise the target functions (maximum entropy or minimum KL) using the L-BFGS algorithm (Byrd et al., 1995) with default settings. We then choose the minimum/maximum achieved over the J starting points in order to improve the chances of achieving a global optimum.

A.2.2 Markov chain Monte Carlo *via* Stan

Most of the posterior distributions discussed in this paper cannot be computed in closed-form and therefore we resort to Hamiltonian Monte Carlo (Neal et al., 2011), in particular the No-U-Turn (NUTS) dynamic implementations available in Stan (Hoffman and Gelman, 2014; Betancourt, 2017).

For most computations, we employed four independent chains of 4000 iterations each with the first 2000 discarded as warm-up/burn-in. Some models presented more challenging target distributions and we increased the number of iterations to 10000. For all results reported, Monte Carlo error (MCSE) was well below 1% the posterior standard deviation for all parameters, allowing for accurate computation of the relevant expectations. In order to cope with challenging posterior geometry and ensure accurate computation, we used an target acceptance probability of 0.99 (`adapt_delta = 0.99`, in Stan parlance) and up to 2^{15} leapfrog steps (`max_treedepth = 15`). All potential scale reduction factors ($\hat{R}$, Gelman and Rubin (1992)) were below 1.01, indicating no convergence problems.

A.2.3 Sampling-importance-resampling

In the context of Bayesian melding, while it is possible to employ HMC, it is sometimes preferable to employ custom algorithms that can better deal with the constraints imposed by the deterministic model. In the original paper, (Poole and Raftery, 2000, sec. 3.4) propose a sampling-importance-resampling (SpIR) algorithm to sample from the posterior in (13), which we extend here to in order to accommodate varying weights.

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

4. Form the importance weights

$$w_i = t\left(\boldsymbol{\alpha}^{(i)}\right)\left(\frac{q_2(M(\theta^{(i)}))}{q_1^{\star}(M(\theta^{(i)}))}\right)^{1-\boldsymbol{\alpha}^{(i)}} L_1(\theta^{(i)})L_2(M(\theta^{(i)})), \tag{42}$$

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

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

The quadrature-based normalisation in step in 4 can be replaced with an importance sampling or MCMC estimate when the dimension of either ϕ or θ is large, but this is not explored here.

A.3 Bowhead population growth model: details

For convenience, here we will describe the priors and likelihoods used by Poole and Raftery (2000) in their analysis of the bowhead whale population model, as well as some of our modelling choices. For P_0 only a shifted gamma prior is available, i.e.,

$$q(P_0) = \frac{b_{P_0}^{a_{P_0}}}{\Gamma(a_{P_0})}(P_0 - s_{P_0})^{a_{P_0}-1}\exp\left(-b_{P_0}(P_0 - s_{P_0})\right),\ P_0 > s_{P_0},$$

with $s_{P_0} = 6400$, $a_{P_0} = 2.8085$ and $b_{P_0} = 0.0002886$. The maximum sustainable yield rate (MSYR) is assigned Gamma prior with parameters $a_{\text{MSYR}} = 8.2$ and $a_{\text{MSYR}} = 372.7$.

The size of the bowhead population in 1993, P_{1993}, is an output of the model for which there are both a prior and a likelihood. The prior (q_2) is a Gaussian distribution with mean $\mu_{1993} = 7800$ and standard deviation $\sigma_{1993} = 1300$, while the likelihood (L_2) is also a Gaussian distribution but with mean $\mu'_{1993} = 8293$ and standard deviation $\sigma'_{1993} = 626$. For the rate of increase (ROI), Poole and Raftery (2000) use a likelihood that is proportional to $\exp(a+b\times t_8)-1$, with $a = 0.0302$ and $b = 0.0068$ where t_8 is a random variable with Student t distribution with $\nu = 8$ degrees of freedom. This leads to the density

$$L(\text{ROI} \mid \nu, a, b) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{1}{\sqrt{\nu\pi}}\left(1 + ((\log(\text{ROI}+1) - a)/b)^2\right)^{-(\nu+1)/2}\frac{1}{|b(\text{ROI}+1)|}, \text{ROI} > -1.$$

For the Stan implementation, we approximate the distribution induced on P_{1993} by the prior on P_0 and MSYR and transformation in (32), q_1^*, by a normal distribution with mean $\mu_{\text{ind}} = 18137.70$ and standard deviation $\sigma_{\text{ind}} = 6146.85$. This step deserves a bit more consideration. As discussed by Poole and Raftery (2000), q_1^* is very diffuse and likely has heavier tails than a normal distribution. Hence it would make sense also to consider the skew-normal and log-normal families as approximating distributions. On the other hand, we note that approximating q_1^* with a normal distribution allows closed-form computation of the coherised prior $\tilde{q}_\Phi(P_{1993})$. In Figure S1 we show the densities of a normal, skew-normal and log-normal distributions fitted to $100,000$ simulations from q_1^* by maximum likelihood. While the skew-normal provides better fit (AIC: 1150814) we do not feel the difference in fit to the normal (AIC: 1152288) justifies the increased technical overhead of not being able to compute the coherised prior in closed-form. Both distributions provide much superior fit than the log-normal (AIC: 1181753).

Note that the sampling-importance-resampling discussed in Section A.2.3 does not necessitate any parametric approximation, employing a density estimation method instead.

In Figure S2 we show the posterior distributions obtained with different values of α using SpIR.

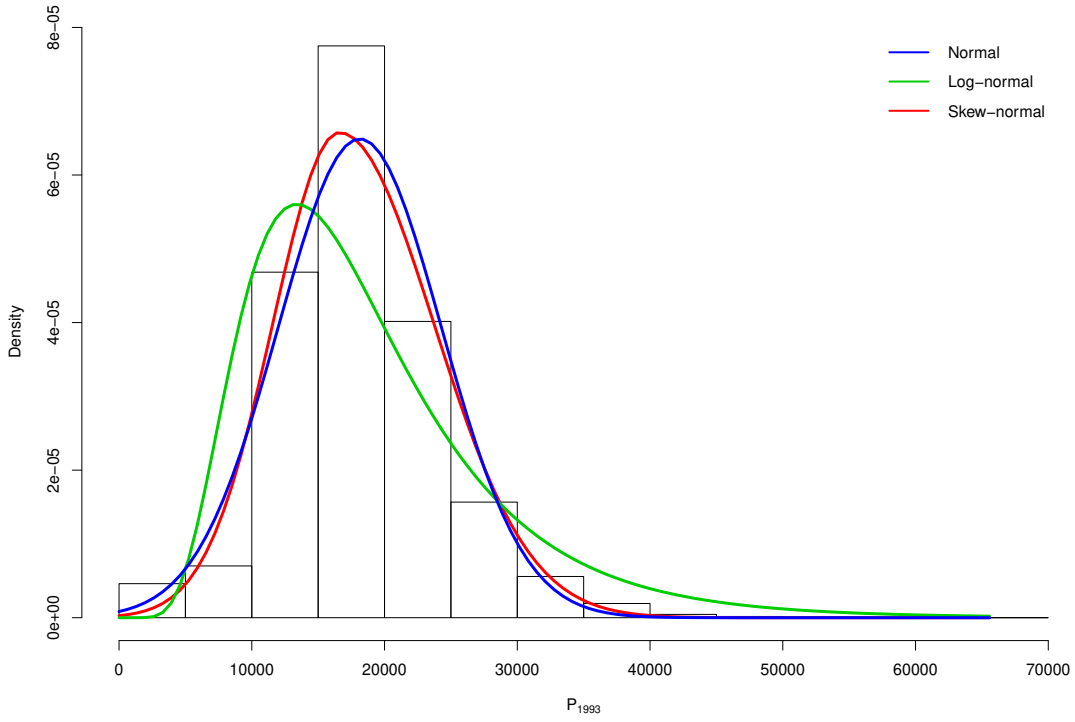


Figure S1: **Induced distribution on P_1993 (q_1^*) and approximating distributions**. We present the histogram of $100,000$ simulations from the prior. Lines show the densities of the three distributions considered.

A.4 Pooling of common distributions

In the main text we give the pooled distributions if one assumes a set of Beta (Section 4.1) or Gaussian (Section 4.2) distributions as the expert opinions. In this section we give further results for the pooling of commonly used distributions.

A.4.1 Gamma

Suppose $K+1$ experts are called upon to elicit prior distributions for a quantity $\lambda \in \mathbb{R}^+$. 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

$$\begin{aligned} \pi(\lambda) &= t(\boldsymbol{\alpha}) \prod_{i=0}^{K} f_i(\lambda; a_i, b_i)^{\alpha_i}, \\ &\propto \prod_{i=0}^{K} \left(\lambda^{a_i - 1} e^{-b_i \lambda} \right)^{\alpha_i}, \\ &\propto \lambda^{a^* - 1} e^{-b^* \lambda}, \end{aligned} \tag{43}$$

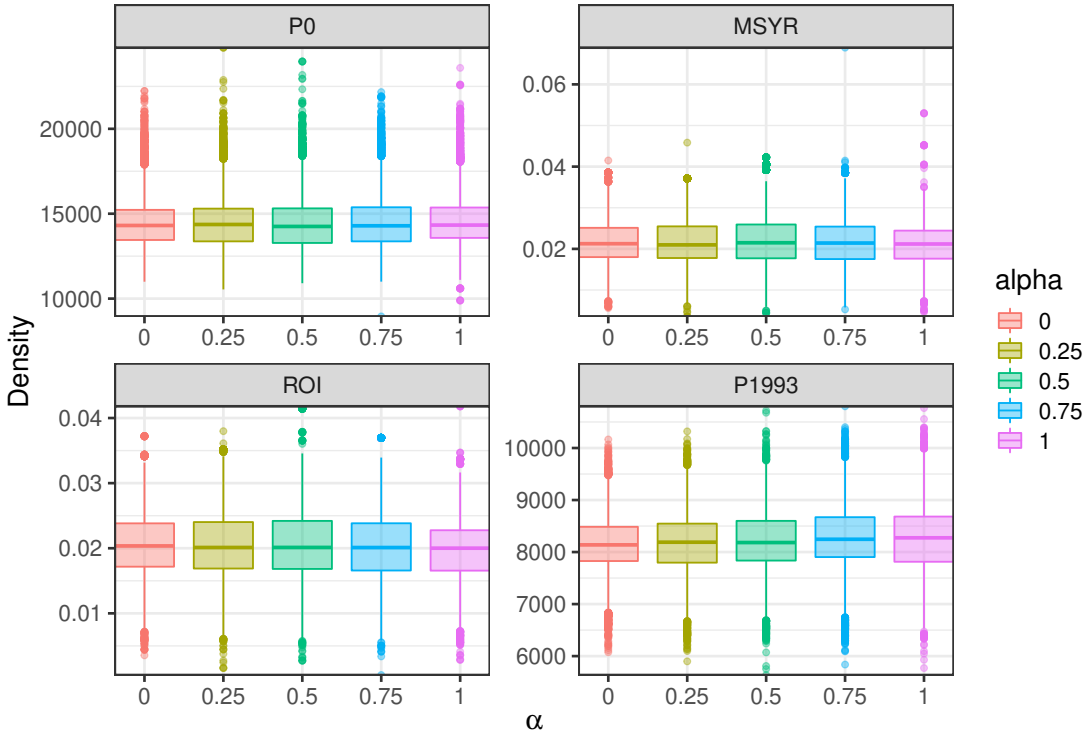


Figure S2: **Sensitivity of posterior inferences to varying the value of α, bowhead population model.**
We show the posterior distributions obtained by SpIR for P_0, MSYR, ROI and P_1993 as we fix α to different values.

where $a^* = \sum_{i=0}^{K} \alpha_i a_i$ and $b^* = \sum_{i=0}^{K} \alpha_i b_i$. Noticing (43) is the kernel of a gamma distribution with parameters a^* and b^*, $H_\pi(\lambda)$ becomes

$$H_\pi(\lambda; \boldsymbol{\alpha}) = a^* - \log b^* + \log \Gamma(a^*) + (1 - a^*)\psi(a^*), \tag{44}$$

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

$$\mathrm{KL}(\pi || f_i) = (a_i - a^*)\psi(a_i) - \log \Gamma(a_i) + \log \Gamma(a^*) + a^* \left(\log \frac{b_i}{b^*} \right) + \frac{a_i}{b_i}(b^* - b_i). \tag{45}$$

A.4.2 Log-normal

Another popular choice for modelling a quantity $\eta \in \mathbb{R}^+$ is the log-normal family. Following the results given in Section 4.2, we know that the pool of log-normal distributions with parameters μ_i and σ_i^2 is a log-normal distribution with parameters $\mu^\star = \frac{\sum_{i=0}^{K} w_i m_i}{\sum_{i=0}^{K} w_i}$ and $\sigma^{\star 2} = [\sum_{i=0}^{K} w_i]^{-1}$, where $w_i = \alpha_i / \sigma_i^2$.

The entropy function is then:

$$\begin{aligned} H_\pi(\eta; \boldsymbol{\alpha}) &= \log_2(e) \log \left(\sigma^\star \exp \left(\mu^\star + \frac{1}{2} \right) \sqrt{2\pi} \right), \\ &= \log_2(e) \left[\log (\sigma^\star) + \mu^\star + \frac{1}{2} + \log(\sqrt{2\pi}) \right]. \end{aligned} \tag{46}$$

The KL divergence evaluates to

$$\mathrm{KL}(\pi||f_i) = \frac{1}{2\sigma_i^2}\left[(\mu^\star - \mu_i)^2 + \sigma^{\star 2} - \sigma_i^2\right] + \log\left(\frac{\sigma_i^2}{\sigma^{\star 2}}\right). \tag{47}$$

A.4.3 Poisson

If the quantity of interest is a count $y = 0, 1, \ldots$, and $\boldsymbol{F}_y$ is a set of Poisson distributions with rate parameters $\boldsymbol{\lambda} = \{\lambda_0, \lambda_1, \ldots, \lambda_K\}$. We have

$$\begin{aligned}
\pi(y) &\propto \prod_{i=0}^{K}\left(\frac{\lambda_i^y}{y!}\right)^{\alpha_i}, \\
\pi(y) &= \frac{\exp(-\lambda^\star)\lambda^{\star y}}{y!}, \text{ with } \lambda^\star = \prod_{i=0}^{K}\lambda_i^{\alpha_i}.
\end{aligned} \tag{48}$$

The entropy of the pooled distribution is

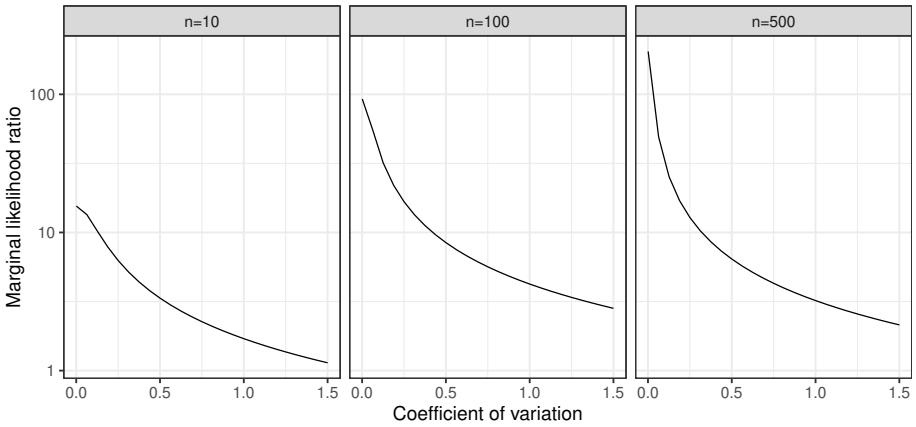
$$H_\pi(y;\boldsymbol{\alpha}) = -\lambda^\star \log\left(\frac{\lambda^\star}{e}\right) + E_\pi\left[\log(k!)\right], \tag{49}$$

where the latter term cannot be evaluated in closed-form, but efficient approximations exist (Evans and Boersma, 1988).

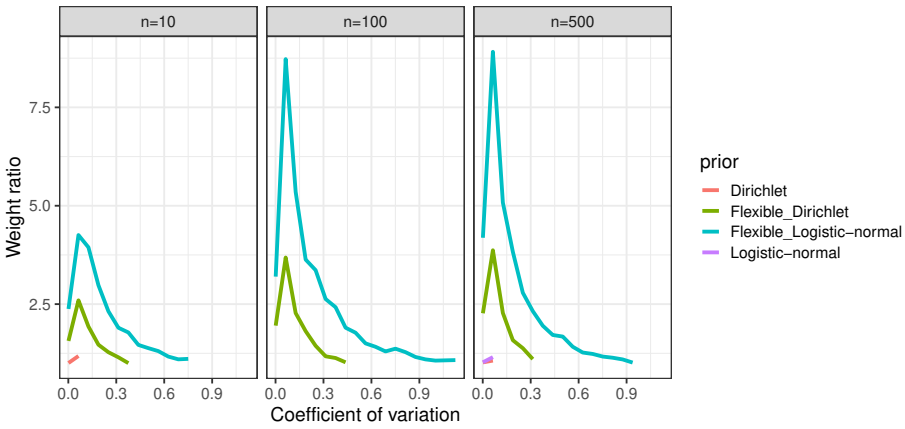
The KL divergence is

$$\begin{aligned}
\mathrm{KL}(\pi||f_i) &= \lambda^\star \log\left(\frac{\lambda^\star}{\lambda_i}\right) + \lambda_i - \lambda^\star, \\
&= \lambda^\star\left[\sum_{k=0}^{K}\alpha_k \log(\lambda_k) - \log(\lambda_i)\right] + \lambda_i - \lambda^\star.
\end{aligned} \tag{50}$$

A.5 Supplementary figures

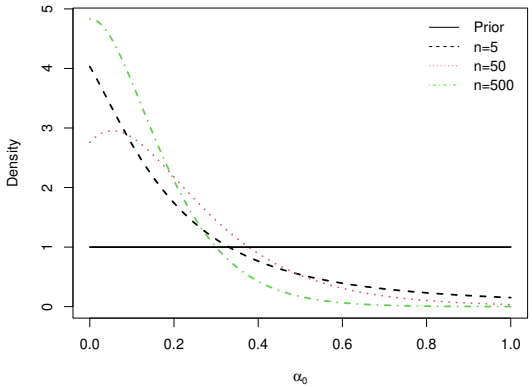


(a) Ratios of marginal likelihoods

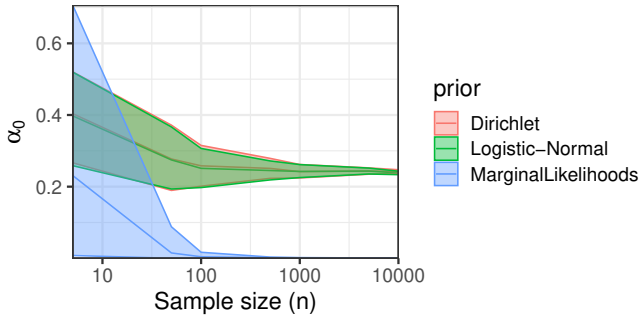


(b) Ratios of posterior weights

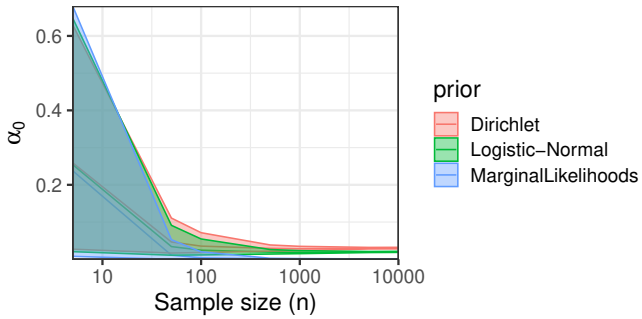
Figure S3: **Marginal likelihood and weight ratios for the simulated situation with one correct expert, various strengths of evidence, Gaussian example**. Panel (a) shows the ratio between the largest and second largest marginal likelihoods (r_l) as the correct expert's coefficient of variation (c_2) changes, while panel (b) shows the ratio between the largest and second largest posterior mean weights (r_w) in the same settings. Vertical tiles show the observed data and colours in panel (b) show the hyperprior on $\boldsymbol{\alpha}$. "Flexible" priors are a Dirichlet(1/10, 1/10, 1/10, 1/10, 1/10) and the corresponding moment-matching logistic-normal. We interrupt the lines for values of c_2 for which expert 2, the correct expert, does not attain the largest posterior weight (see Section 4.3.2).



(a) Marginal posterior of α_0



(b) Posterior concentration, uniform prior



(c) Posterior concentration, Beta$(1/10, 1/10)$ prior

Figure S4: **Posterior concentration for the weights in a two-expert setting, Gaussian conjugate analysis with known variance (σ^2).** In this experiment we set the expert hyper-parameters to $m_0 = 1$ $v_0 = (1/4)^2$, $m_1 = 2$, $v_1 = (1/2)^2$ and the true data-generating parameters at $\mu = 2$ and $\sigma^2 = 1^2$. In panel (a) we show the marginal posterior of α_0 for various sample sizes under a Beta prior for α_0 with parameters $a = b = 1$. Panel (b) shows the posterior mean of α_0 versus sample size under repeated sampling, using 100 simulated data sets per sample size. Bands correspond to the 95% quantiles of the sampling distribution of the posterior mean. In this experiment the prior on α_0 is a uniform prior over $(0, 1)$. For comparison we show the BMA weight for expert 0, computed from the marginal likelihoods. In panel (c) we show the same experiment, but using the "flexible" prior discussed in the main text, corresponding to a Beta$(1/10, 1/10)$ prior on α_0.