Logarithmic pooling of probability distributions: advances and prospects.[1]

Luiz Max F. de Carvalho [lmax.procc@gmail.com], Daniel Villela, Flávio Coelho & Leonardo S. Bastos

School of Applied Mathematics, Getúlio Vargas Foundation (FGV), Rio de Janeiro, Brazil.
Presented at the School of Mathematics and Statistics, the Open Universtity, UK.

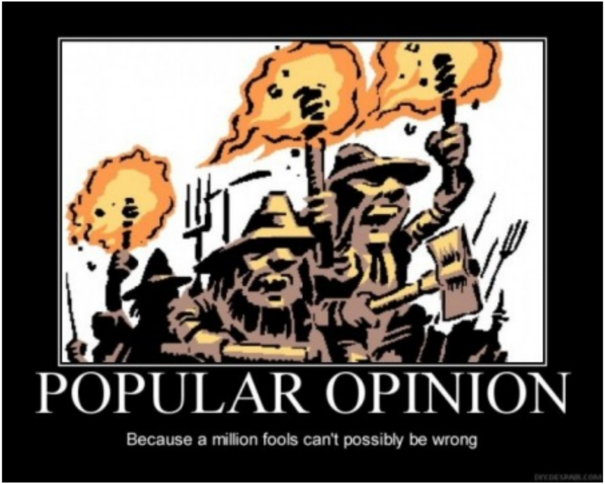
May 25, 2021

[1]This talk is available on-line at https://github.com/maxbiostat/opinion_pooling/blob/master/presentations/open_university_2021/lmcarvalho_OpenUni_2021.pdf

Plan for today

- Part I: paper.
  - Review and problem statement;
  - Hierarchical modelling of the weights;
  - Examples + Discussion.
- Part II: avenues of research.
  - Transformations;
  - Connection with other theories (Bayesian predictive synthesis, BPS);
  - Practical challenges.

Combining (expert) opinions



Let $\mathsf{F}_\theta = \{f_0(\theta), f_1(\theta), \ldots, f_K(\theta)\}$ be the set of ~~prior~~ distributions representing the opinions of $K+1$ experts and let $\boldsymbol{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_K\}$ 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

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

where the normalising term $t(\boldsymbol{\alpha}) = \int_\Theta \prod_{i=0}^K f_i(\theta)^{\alpha_i} d\theta$ is guaranteed to exist for all proper f_i. We simplify the proof given by Genest et al. (1986) by using Hölder's inequality. This operator enjoys a number of desirable properties such as external Bayesianity (Genest et al., 1986), relative propensity consistency (Genest et al., 1984) and log-concavity (Carvalho et al., 2021).

Logarithmic pooling – Properties

Property 1

External Bayesianity **(*Genest et al., 1984*)***.* *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* $p(\theta \mid x) \propto l(x \mid \theta)\pi(\theta)$.

Property 2

Log-concavity **(*Carvalho et al., 2021*)***.* *Let* F_θ *be a set of log-concave distributions, i.e., each* f_i *can be written as* $f_i(\theta) \propto e^{\nu_i(\theta)}$*, where* $\nu_i(\cdot)$ *is a concave function. Then* $\pi(\theta \mid \boldsymbol{\alpha})$ *is also log-concave. Moreover, LP is the* ***only*** *operator that* ***always*** *preserves log-concavity.*

Logarithmic pooling – more properties

Property 3

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

$$U(\boldsymbol{\xi}) := \left(\frac{f_0(a)}{f_0(b)}, \frac{f_1(a)}{f_1(b)}, \ldots, \frac{f_K(a)}{f_K(b)}\right) \text{ and } \tag{2}$$

$$V(\boldsymbol{\xi}) := \frac{\mathcal{T}_{\boldsymbol{F}_X}(a)}{\mathcal{T}_{\boldsymbol{F}_X}(b)}. \tag{3}$$

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

$$U(\boldsymbol{\xi}_1) \geq U(\boldsymbol{\xi}_2) \implies V(\boldsymbol{\xi}_1) \geq V(\boldsymbol{\xi}_2), \tag{4}$$

for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$.

- All of these properties, especially 1 and 3 are **unique** to logarithmic pooling.

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;
- In practice, reliabilities are hard to determine;

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;
- In practice, reliabilities are hard to determine;
- As we will see, combining probability distributions in particular can lead to counterintuitive results.

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;
- In practice, reliabilities are hard to determine;
- As we will see, combining probability distributions in particular can lead to counterintuitive results.

Here we will consider three ways of constructing the weights "objectively":

➔ Maximise the entropy of π;

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;
- In practice, reliabilities are hard to determine;
- As we will see, combining probability distributions in particular can lead to counterintuitive results.

Here we will consider three ways of constructing the weights "objectively":

➔ Maximise the entropy of π;

➔ Minimise the Kullback-Leibler divergence between π and each f_i;

Weights are crucial

The weights α are key to the shape and properties of the combined (pooled) distribution.

- In theory, the weights should reflect the *reliabilities* of the experts/information sources;
- In practice, reliabilities are hard to determine;
- As we will see, combining probability distributions in particular can lead to counterintuitive results.

Here we will consider three ways of constructing the weights "objectively":

➔ Maximise the entropy of π;

➔ Minimise the Kullback-Leibler divergence between π and each f_i;

➔ Place a probability measure over $\mathcal{A} \subset [0, 1]^{K+1}$.

Maximise the entropy of $\pi(\theta)$

- If there is no information about the reliabilities of the experts one might want to construct $\boldsymbol{\alpha}$ so as to maximise entropy of the resulting distribution:

$$H_\pi(\theta) = -\int_\Theta \pi(\theta) \ln \pi(\theta) d\theta$$

$$H_\pi(\theta; \boldsymbol{\alpha}) = -\sum_{i=0}^{K} \alpha_i E_\pi[\log f_i] - \log t(\boldsymbol{\alpha}).$$

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

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

- Caveats: (i) is not guaranteed to yield an unique solution; (ii) is rather prone to yield "degenerate" (trivial) solutions.

Minimise KL divergence between $\pi(\theta)$ and the f_i's

- What if we want to minimise conflict between the consensus and each individual opinion?
- Let $d_i = \text{KL}(f_i || \pi)$ and let $L(\boldsymbol{\alpha})$ be a loss function such that

$$
\begin{aligned}
L(\boldsymbol{\alpha}) &= \sum_{i=0}^{K} d_i \\
&= -(K+1)\log t(\boldsymbol{\alpha}) - (K+1)\sum_{i=0}^{K} \alpha_i \mathbb{E}_i[\log f_i] + \sum_{i=0}^{K} \mathbb{E}_i\left[\log f_i\right],
\end{aligned}
$$

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

- Contrary to the maximum entropy problem, the loss function is convex, thus there is a unique solution (Rufo et al., 2012).

Place a prior on the weights

- An appealing alternative is to place a (hyper) prior on the weights ($\boldsymbol{\alpha}$);
- Two approaches
  (a) Dirichlet prior:

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

  (b) logistic-normal:

$$\pi_A(\boldsymbol{\alpha} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} \frac{1}{\prod_{i=0}^{K} \alpha_i} \exp\left((\boldsymbol{\eta} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}) \right),$$

$$\boldsymbol{\eta} := \log\left(\frac{\boldsymbol{\alpha}_{-K}}{\alpha_K} \right).$$

- Advantage: accomodates uncertainty in natural way, and is very flexible;
- Caveat(s): may yield inconsistent results and hardly ever allows for analytical solutions for the marginal prior $g(\theta) = \int_{\mathcal{A}} \pi(\theta \mid \boldsymbol{\alpha}) d\Pi_A$.

Priors on the weights: details

- Match the first two moments of the Logistic-normal to the Dirichlet (Aitchison and Shen, 1980):

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

  where $\psi(\cdot)$ is the digamma function, and $\psi'(\cdot)$ is the trigamma function.
- Exploit a non-centering trick to sample from the logistic normal *via* Cholesky decomposition of Σ;
- We explore two sets of hyperparameters: $\boldsymbol{X} = \{1, 1, \ldots, 1\}$ and $\boldsymbol{X}' = \boldsymbol{X}/10$ ("flexible" henceforth);

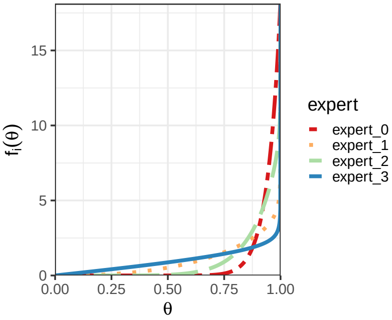
Application: survival probabilities (reliability)

- Savchuk and Martz (1994) consider an example in which four experts are required supply prior information about the survival probability of a certain unit for which there have been $y = 9$ successes out of $n = 10$ trials;
- $Y \sim \textit{Bernoulli}(\theta)$ and

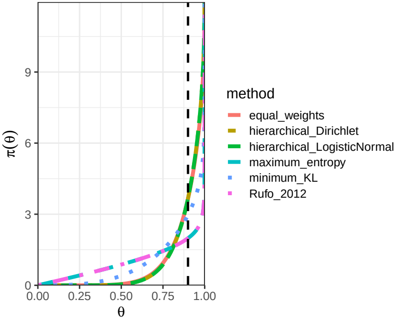
$$f_i(\theta; a_i, b_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i b_i)} \theta^{a_i - 1} (1 - \theta)^{b_i - 1}$$

- Allows for simple expressions for the entropy and KL divergence – $\pi(\theta \mid \boldsymbol{\alpha})$ is also Beta – and efficient sampling from the hyperpriors;
- For this example, we can evaluate performance using integrated (marginal) likelihoods, a.k.a., prior evidence.

Survival probabilities: results



(a) Expert priors



(b) Pooled priors

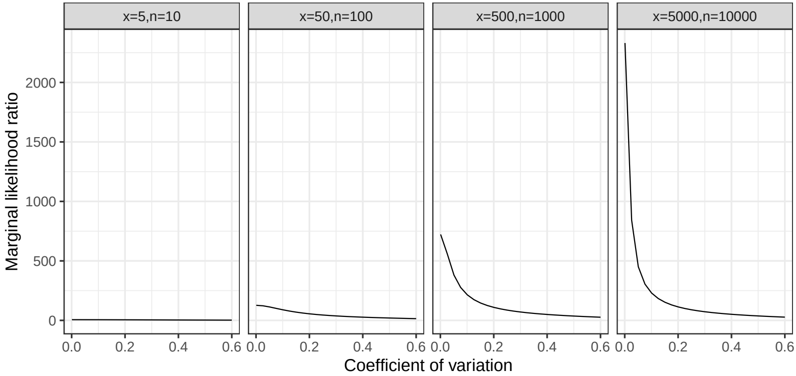
Survival probabilities: results II

Method	α_0	α_1	α_2	α_3
Maximum entropy	0.00	0.00	0.00	1.00
Minimum KL divergence	0.04	0.96	0.00	0.00
Rufo et al. (2012)	0.00	0.00	0.00	1.00
Dirichlet prior	0.26	0.24	**0.27**	0.23
Logistic-normal prior	0.27	0.24	**0.31**	0.18

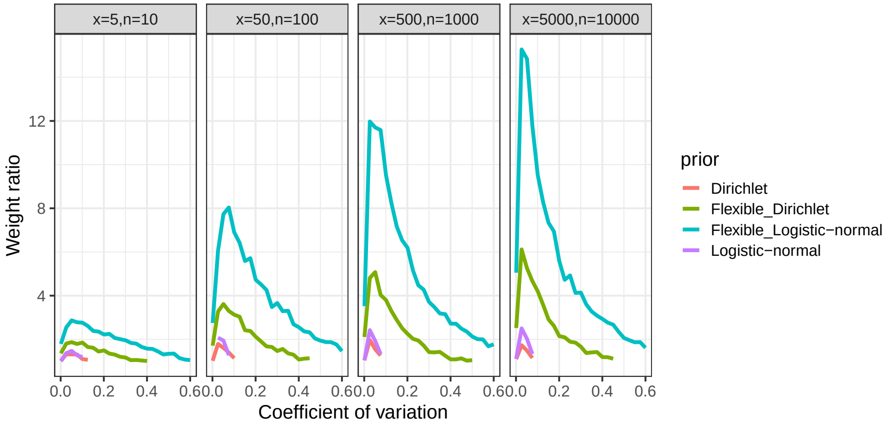
Expert priors		Pooled priors	
Expert 0	0.237	Equal weights	0.254
Expert 1	0.211	Maximum entropy	0.163
Expert 2	**0.256**	Minimum KL	0.223
Expert 3	0.163	Hierarchical prior (Dirichlet/logistic-normal)	**0.255**

Simulated example: can we reliably learn the weights?

Setup: Five experts elicit Beta priors on a quantity p. Data will be $x/n = 5/10$. Only expert 2 (let's call her Mãe Diná) gives a reasonable prior with mean $\mu_2 = 0.50$ and coefficient of variation c_2.



Simulated example: performance of hierarchical priors



Simulated example: explaining the weirdness

- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;

Simulated example: explaining the weirdness

- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;
- This setup leads to $\boldsymbol{a} = \{89.9, 79.8, 12.0, 19.2, 9.1\}$ and $\boldsymbol{b} = \{809.1, 319.2, 12.0, 4.8, 1.01\}$;

Simulated example: explaining the weirdness

- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;
- This setup leads to $\boldsymbol{a} = \{89.9, 79.8, 12.0, 19.2, 9.1\}$ and $\boldsymbol{b} = \{809.1, 319.2, 12.0, 4.8, 1.01\}$;
- If the data are $x = 5$ and $n = 10$, computing marginal likelihoods and normalising would lead to weights $\boldsymbol{\alpha}'' = \{0.006, 0.095, 0.710, 0.142, 0.048\}$;

Simulated example: explaining the weirdness

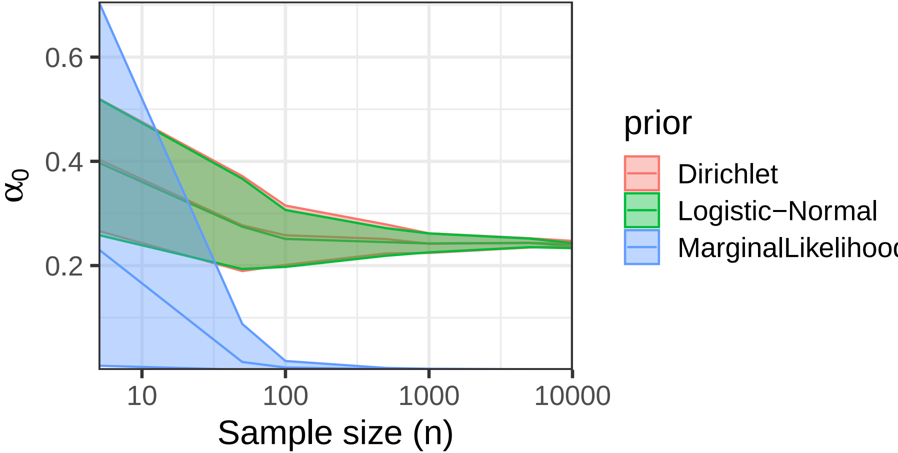
- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;
- This setup leads to $\boldsymbol{a} = \{89.9, 79.8, 12.0, 19.2, 9.1\}$ and $\boldsymbol{b} = \{809.1, 319.2, 12.0, 4.8, 1.01\}$;
- If the data are $x = 5$ and $n = 10$, computing marginal likelihoods and normalising would lead to weights $\boldsymbol{\alpha}'' = \{0.006, 0.095, 0.710, 0.142, 0.048\}$;
- However, by calculating $a^{\star\star} = \sum_{i=0}^{K} \alpha_i'' a_i = 19.75$ and $b^{\star\star} = \sum_{i=0}^{K} \alpha_i'' b_i = 44.00$, we obtain a pooled prior with $\mathbb{E}_\pi[p] = 0.31$, far off the "optimal" $1/2$;

Simulated example: explaining the weirdness

- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;
- This setup leads to $\boldsymbol{a} = \{89.9, 79.8, 12.0, 19.2, 9.1\}$ and $\boldsymbol{b} = \{809.1, 319.2, 12.0, 4.8, 1.01\}$;
- If the data are $x = 5$ and $n = 10$, computing marginal likelihoods and normalising would lead to weights $\boldsymbol{\alpha}'' = \{0.006, 0.095, 0.710, 0.142, 0.048\}$;
- However, by calculating $a^{\star\star} = \sum_{i=0}^{K} \alpha_i'' a_i = 19.75$ and $b^{\star\star} = \sum_{i=0}^{K} \alpha_i'' b_i = 44.00$, we obtain a pooled prior with $\mathbb{E}_\pi[p] = 0.31$, far off the "optimal" 1/2;
- If the data were, say, $x = 50, n = 100$, then one would obtain a pooled prior for which $\mathbb{E}_\pi[p] = 0.51$.

Simulated example: explaining the weirdness

- Let $c_2 = 0.2$ and $c_j = 0.1$ for all $j \neq 2$, with $\boldsymbol{\mu} = \{0.1, 0.2, 0.5, 0.8, 0.9\}$;
- This setup leads to $\boldsymbol{a} = \{89.9, 79.8, 12.0, 19.2, 9.1\}$ and $\boldsymbol{b} = \{809.1, 319.2, 12.0, 4.8, 1.01\}$;
- If the data are $x = 5$ and $n = 10$, computing marginal likelihoods and normalising would lead to weights $\boldsymbol{\alpha}'' = \{0.006, 0.095, 0.710, 0.142, 0.048\}$;
- However, by calculating $a^{\star\star} = \sum_{i=0}^{K} \alpha_i'' a_i = 19.75$ and $b^{\star\star} = \sum_{i=0}^{K} \alpha_i'' b_i = 44.00$, we obtain a pooled prior with $\mathbb{E}_\pi[p] = 0.31$, far off the "optimal" $1/2$;
- If the data were, say, $x = 50$, $n = 100$, then one would obtain a pooled prior for which $\mathbb{E}_\pi[p] = 0.51$.
- Now let $c_2 = 0.001$. Then $a_2 = b_2 = 499999.5$. Can you see the problem?



0.6
0.4
0.2
α_0
10
100
1000
10000
Sample size (n)
prior
Dirichlet
Logistic−Normal
MarginalLikelihood

Bayesian melding

Suppose we have deterministic model M with inputs $\theta \in \Theta \subseteq \mathbb{R}^p$ and outputs $\phi \in \Phi \subseteq \mathbb{R}^q$, such that $\phi = M(\theta)$. We have the combined prior on the outputs:

$$\tilde{q}_{\Phi}(\phi) \propto q_1^*(\phi)^{\alpha} q_2(\phi)^{1-\alpha}, \tag{5}$$

where $q_1^*()$ is the **induced** and q_2 is "natural" prior on ϕ. The prior in (5) 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

$$p_{\Theta}(\theta \mid \mathbf{y}, \alpha) \propto \tilde{q}_{\Theta}(\theta) L_1(\theta) L_2(M(\theta)) \pi_A(\alpha). \tag{6}$$

Application: Influenza in a boarding school

In 1978, 512 out of 763 lads got came down with the flu. We model the spread using a standard SIR model

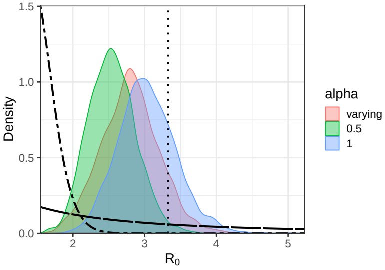
$$\begin{aligned}\frac{dS}{dt} &= -\beta S I,\\ \frac{dI}{dt} &= \beta S I - \gamma I,\\ \frac{dR}{dt} &= \gamma I,\end{aligned}$$

where $S(t) + I(t) + R(t) = N \, \forall t$, β is the transmission (infection) rate and γ is the recovery rate. The basic reproductive number is

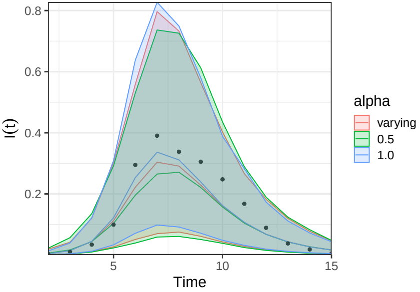
$$R_0 = \frac{\beta N}{\gamma}. \tag{7}$$

We choose $\beta, \gamma \sim$ log-normal$(0, 1)$ (q_1) and $R_0 \sim$ log-normal(μ_2, σ_2) (q_2) such that R_0 has a mean of 1.5 and a standard deviation of 0.25 (q_2) which is informed by seasonal flu.

SIR model: results



(c) R_0



(d) Predictions

Posterior for α: **0.77** (0.18–0.99).

Partial Sum up

- Optimality criteria often give weird results;

Partial Sum up

- Optimality criteria often give weird results;
- It is possible to learn about the weights from data, for some configurations of the opinions $\boldsymbol{F}_X$ (and the data $\boldsymbol{y}$);

Partial Sum up

- Optimality criteria often give weird results;
- It is possible to learn about the weights from data, for some configurations of the opinions $\boldsymbol{F}_X$ (and the data $\boldsymbol{y}$);
- Interpretation of the weights is not always straightforward;

Partial Sum up

- Optimality criteria often give weird results;
- It is possible to learn about the weights from data, for some configurations of the opinions $\boldsymbol{F}_X$ (and the data $\boldsymbol{y}$);
- Interpretation of the weights is not always straightforward;
- In Bayesian melding, letting the pooling weight α vary can protect against prior misspecification.

Part II: “Cool. What else?”

Induce-then-pool or pool-then-induce?

Question: what to do when $M : \Theta \to \Phi$ is non-invertible? We may want to gain insight about ϕ, even though we only have expert opinions about θ.

➔ If we apply $M(\cdot)$ to each component of F_θ, we get a set induced distributions $\boldsymbol{G}_\phi$, which are then pooled to get $\pi_P(\phi)$ [**induce-then-pool**];

Induce-then-pool or pool-then-induce?

Question: what to do when $M : \Theta \to \Phi$ is non-invertible? We may want to gain insight about ϕ, even though we only have expert opinions about θ.

- If we apply $M(\cdot)$ to each component of $\mathbf{F}_\theta$, we get a set induced distributions $\boldsymbol{G}_\phi$, which are then pooled to get $\pi_P(\phi)$ [**induce-then-pool**];
- Alternatively, we can combine the $f_i(\theta)$ to obtain $\pi_T(\theta)$ and then transform to get a distribution $\pi'_P(\phi)$ [**pool-then-induce**];

Induce-then-pool or pool-then-induce?

Question: what to do when $M : \Theta \to \Phi$ is non-invertible? We may want to gain insight about ϕ, even though we only have expert opinions about θ.

- ➔ If we apply $M(\cdot)$ to each component of F_θ, we get a set induced distributions $\boldsymbol{G}_\phi$, which are then pooled to get $\pi_P(\phi)$ [**induce-then-pool**];
- ➔ Alternatively, we can combine the $f_i(\theta)$ to obtain $\pi_T(\theta)$ and then transform to get a distribution $\pi_P'(\phi)$ [**pool-then-induce**];
- **Remark:** if $M(\cdot)$ is invertible, $\pi_P(\phi) \equiv \pi_P'(\phi)$.

Induce-then-pool or pool-then-induce?

Question: what to do when $M : \Theta \to \Phi$ is non-invertible? We may want to gain insight about ϕ, even though we only have expert opinions about θ.

➔ If we apply $M(\cdot)$ to each component of F_θ, we get a set induced distributions $\boldsymbol{G}_\phi$, which are then pooled to get $\pi_P(\phi)$ [**induce-then-pool**];

➔ Alternatively, we can combine the $f_i(\theta)$ to obtain $\pi_T(\theta)$ and then transform to get a distribution $\pi'_P(\phi)$ [**pool-then-induce**];

- **Remark:** if $M(\cdot)$ is invertible, $\pi_P(\phi) \equiv \pi'_P(\phi)$.
- When M is non-invertible, things get complicated, as we shall see.

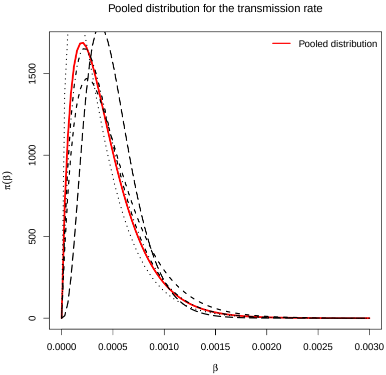
Recall that $\theta = \{\beta, \gamma\}$ and $M(\theta) = R_0$. Suppose $p(\beta, \gamma) = p(\beta)p(\gamma)$.

Useful result:

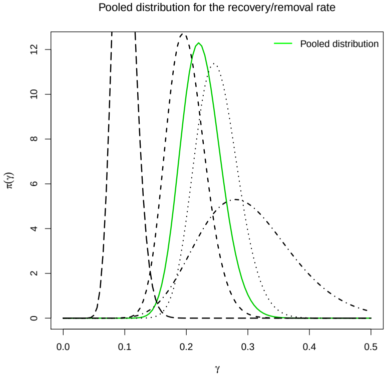
If $\beta \sim \text{Gamma}(k_\beta, t_\beta)$ and $\gamma \sim \text{Gamma}(k_\gamma, t_\gamma)$, then

$$f_{R_0}(r \mid k_\beta, t_\beta, k_\gamma, t_\gamma, N) = \frac{(N t_\beta t_\gamma)^{k1+k2}}{\mathcal{B}(k_\beta, k_\gamma)(N t_\beta)^{k_\beta} t_\gamma^{k_\gamma}} R_0^{k_\beta - 1}(t_\gamma R_0 + N t_\beta)^{-(k_\beta + k_\gamma)}$$

where $\mathcal{B}(a, b) = \Gamma(a+b)/\Gamma(a)\Gamma(b)$ is the Beta function.


Pooled distribution for the transmission rate
Pooled distribution
π(β)
0
500
1000
1500
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
β

(e)


Pooled distribution for the recovery/removal rate
Pooled distribution
π(γ)
0
2
4
6
8
10
12
0.0
0.1
0.2
0.3
0.4
0.5
γ

(f)

- Pool:

$$\pi(\beta) = Gamma(t_1^*, k_1^*)$$
$$\pi(\gamma) = Gamma(t_2^*, k_2^*)$$

  where $t^* = \sum_{i=0}^{K} \alpha_i t_i$ and $k^* = \sum_{i=0}^{K} \alpha_i k_i$. Then

- Induce:

$$\pi(R_0) \propto {R_0}^{k_1 - 1}(t_2^* R_0 + N t_1^*)^{-(k_1^* + k_2^*)}$$

- Nice!

Induce-then-pool

- Induce (transform) each distribution (Gamma ratio):

$$g_i(R_0) \propto {R_0}^{k_1-1}(t_2 R_0 + N t_1)^{-(k_1+k_2)}$$

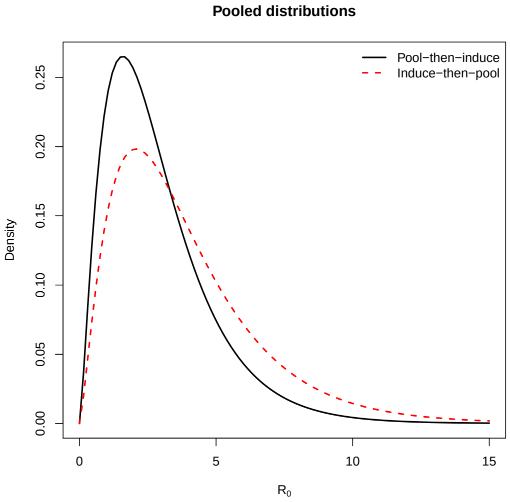
  then
- Pool:

$$\pi'(R_0) \propto \prod_{i=0}^{K} g_i(R_0)^{\alpha_i}$$

- Ugly!

Question: can we derive a formal, measure-theoretical framework to further our understanding?

Pool-then-induce *vs* Induce-then-pool



But if tis' piecewise invertible...

FGV EMAp

Remark 1

It is possible to have $\pi_\Phi \equiv \pi'_\Phi$ even when M is not invertible.

Proof.

By an explict example. Let $\theta \sim \operatorname{normal}(0, \sigma^2)$ and let $M(\theta) = \theta^2$. If we define $\Omega(\phi) := \{x : M(x) = \phi\}$ then clearly $\Omega(\phi) = \{\omega_0, \omega_1\} = \{-\sqrt{\phi}, \sqrt{\phi}\}$ and hence

$$
\begin{aligned}
g_i(\phi) &= \frac{f_i(\omega_0)}{|2\omega_0|} + \frac{f_i(\omega_1)}{|2\omega_1|}, \\
&= \frac{f_i(\sqrt{\phi})}{\sqrt{\phi}} = \frac{1}{\sqrt{2\pi v_i \phi}} \exp\left(-\frac{\phi}{2v_i}\right),
\end{aligned}
$$

where the second line follows by using the symmetry of f_i around zero. Rest of the proof follows analogously to the arguments in Carvalho et al. (2021) for the pool of Gaussians. □

Bayesian predictive synthesis

Bayesian predictive synthesis (BPS) is a general framework for combining models/forecasts (see McAlinn et al. (2018)). Take $\mathcal{H} = \{h_1, \ldots, h_J\}$ a set of densities on Y and write:

$$\pi(y \mid \mathcal{H}) = \int \alpha(y \mid \boldsymbol{X}) \prod_{j=1}^{J} h_j(x_j)\, dx_j.$$

Taking $\alpha(y \mid \boldsymbol{X}) = \sum_{j=1}^{J} w_j \delta_{x_j}(y)$, for instance, recovers the usual Bayesian model averaging (BMA): $\pi(y \mid \mathcal{H}) = \sum_{j=1}^{J} w_j h_j(y)$.

Remark 2 (BPS recovers LP in a toy example)

Assume that

- A_1: *the h_j are Normal densities with parameters m_j and v_j;*
- A_2: *the parameters m_j and v_j **do not** depend on any x_i $i \neq j$.*

Then BPS recovers the Normal pool if we pick

$$\alpha(y \mid \mathbf{x}) = \prod_{j=1}^{J} g_j(y \mid x_j), \tag{8}$$

with

$$g_j(y \mid x_j) = \frac{1}{\sqrt{2\pi\,(1/w_j - 1)\,v_j}} \exp\left(-\frac{(y - x_j)^2}{2\,(1/w_j - 1)\,v_j}\right).$$

Question: can we find a general solution that does **not** depend on $\mathcal{H}$?

Practical hurdles

Suppose one has fitted K models to a set of data and thus obtained M samples for each of the K posterior predictive distributions. One is only able to evaluate the unnormalised posterior densities $f_i^{\star}(\tilde{x} \mid x)$ and the Z_i are unknown. We want to sample from

$$\pi(\tilde{x} \mid x, \boldsymbol{\alpha}) = t(\boldsymbol{\alpha}) \prod_{i=1}^{K} f_i(\tilde{x} \mid x)^{\alpha_i},$$

using an $M \times K$ matrix of samples

$$\begin{aligned} \tilde{\boldsymbol{X}} = \; & \tilde{X}_1^{(1)}, \ldots, \tilde{X}_1^{(M)} \\ & \tilde{X}_2^{(1)}, \ldots, \tilde{X}_2^{(M)} \\ & \qquad \vdots \\ & \tilde{X}_K^{(1)}, \ldots, \tilde{X}_K^{(M)} \end{aligned}$$

and a fixed vector of weights $\boldsymbol{\alpha}$.

Question: can we devise an (efficient!) importance sampling scheme to re-use the samples?

$\boldsymbol{\tau}_h$: Log-linear mixtures have many desirable properties;

Take home

$\boldsymbol{\tau}_h$: Log-linear mixtures have many desirable properties;

$\boldsymbol{\tau}_h$: It is possible to learn the weights from data, under certain constraints;

Take home

- $\boldsymbol{\tau}_h$: Log-linear mixtures have many desirable properties;
- $\boldsymbol{\tau}_h$: It is possible to learn the weights from data, under certain constraints;
- $\boldsymbol{\tau}_h$: Letting the weights vary can lead to more flexible prior modelling and better posterior inferences;

Take home

- τ_h : Log-linear mixtures have many desirable properties;
- τ_h : It is possible to learn the weights from data, under certain constraints;
- τ_h : Letting the weights vary can lead to more flexible prior modelling and better posterior inferences;
- τ_h : Further work on both theory and methods is needed in order to make logarithmic pooling widely applicable in Statistics.

Thank you!

- Thank you very much for your attention!
- The authors would like to thank Professors Adrian Raftery (Washington) and Mike West (Duke) for helpful suggestions. DAMV was supported in part by Capes under Capes/Cofecub project (N. 833/15). FCC is grateful to Fundação Getulio Vargas for funding during this project.
- All the necessary code and data are publicly available at https://github.com/maxbiostat/opinion_pooling

References

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

Carvalho, L. M., Villela, D. V., Coelho, F. C., and Bastos, L. S. (2021). Combining probability distributions: Extending the logarithmic pooling approach. *Submitted*.

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.

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.

Rufo, M., Martin, J., Pérez, C., et al. (2012). Log-linear pool to combine prior distributions: A suggestion for a calibration-based approach. *Bayesian Analysis*, 7(2):411–438.

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.