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

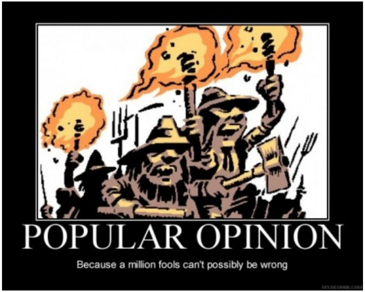
Luiz Max F. de Carvalho [lmax.procc@gmail.com], Daniel Villela, Flavio Coelho & Leonardo S. Bastos [lsbastos@fiocruz.br]

Scientific Computing Programme (PROCC), Oswaldo Cruz Foundation, Fiocruz, Brazil.
XIII Brazilian Bayesian Statistics Meeting (EBEB 2016), Belo Horizonte, Brazil.

February 23, 2016

Logarithmic pooling – Motivation

- Obtain important insights on consensus belief formation and group decision making (Genest and Zidek, 1986);



- Applications in a range of fields, from infectious disease modelling (Coelho and Codeço, 2009) and wildlife conservation (Poole and Raftery, 2000) to engineering (Savchuk and Martz, 1994);
- BUT how to give each expert/information source a weight without being (totally) arbitrary?

Logarithmic pooling – Definition & Notation

Let $\mathbf{F}_\theta = \{f_0(\theta), f_1(\theta), \ldots, f_K(\theta)\}$ 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

$$\pi(\theta) = t(\boldsymbol{\alpha}) \prod_{i=0}^{K} f_i(\theta)^{\alpha_i} \tag{1}$$

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

- Enjoys rather desirable properties, such as *external Bayesianity* (Genest and Zidek, 1986);
- Poole and Raftery (2000) prove that $t(\boldsymbol{\alpha})$ is always finite for the case $K = 1$ (2 experts/priors), which we extend for any finite K. See Theorem 1 in http://arxiv.org/pdf/1502.04206v1.pdf for a simple proof.

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

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

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

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

- 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 “trivial” solutions.

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

- 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

$$L(\boldsymbol{\alpha}) = \sum_{i=0}^{K} d_i \tag{2}$$

$$= -K \ln t(\boldsymbol{\alpha}) + \sum_{i=0}^{K} \sum_{j \neq i}^{K} \alpha_j \text{KL}(f_i || f_j) \tag{3}$$

$$\hat{\boldsymbol{\alpha}} := \arg\min L(\boldsymbol{\alpha}) \tag{4}$$

- Contrary to the above, 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 options:
  (a) Dirichlet prior:
$$\pi(\boldsymbol{\alpha}) = \frac{1}{\mathcal{B}(\mathbf{X})} \prod_{i=0}^{K} \alpha_i^{x_i - 1}$$
  (b) logistic-normal:
$$\alpha_i = \frac{e^{m_i}}{\sum_{i=0}^{K} e^{m_i}},\ m_i \sim N(\mu_i, \sigma_i^2)$$
- Advantage: accomodates uncertainty in natural way, and is very flexible;
- Caveat(s): may yield inconsistent results and hardly ever allows for analytical solutions.

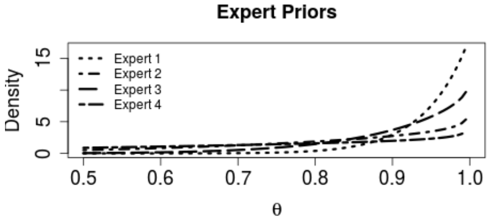
Application: binomial probabilities

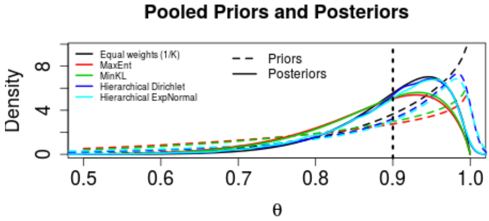
- $Y \sim \mathit{Bernoulli}(\theta)$ and

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

- Allows for simple expressions for the entropy and KL divergence [$\pi(\theta; \boldsymbol{\alpha})$ is also Beta], and efficient sampling from the hyperpriors;
- 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;
- We propose to evaluate performance using integrated (marginal) likelihoods, a.k.a., prior evidence.

~~very~~ Preliminary results





~~very~~ Preliminary results II

Table : 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

Table : Integrated likelihoods 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

Partial Sum up

- Our results are not yet decisive regarding which method is better;
- The Dirichlet approach seems the most natural from a Bayesian perspective, but prior sensitivity is currently unknown;

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

- Let $\theta \in \Theta \subseteq \mathbb{R}^p$ and $y \in \mathcal{Y} \subseteq \mathbb{R}^p$ and define the model (transformation) as $M : \Theta \to \mathcal{Y}$.
- Finally recall $\mathbf{F}_\theta$ is a set of K distributions on θ.
- We may want to gain insight into y, even though we only have expert opinions on θ. If we apply $M(\cdot)$ to each component of $\mathbf{F}_\theta$, we get an **induced** distribution.
- Theorem: if $M(\cdot)$ is invertible, the order in which one pools or induces (transforms) the distributions does not matter.
- This is not always the case, though, as we shall see.

Dynamic model example

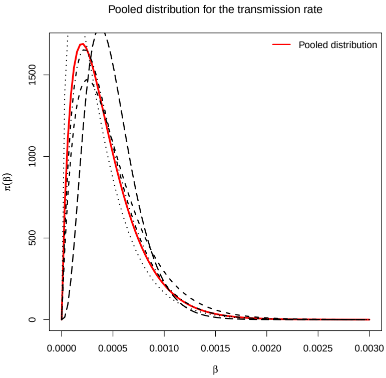
- Susceptible-Infectious-Removed (SIR) epidemic model:

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

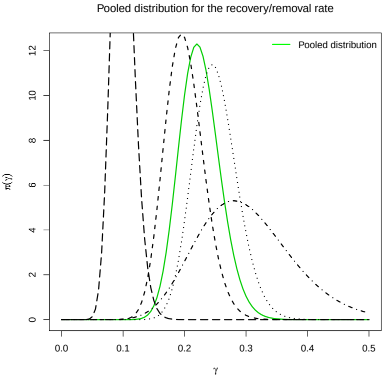
  where $S(t) + I(t) + R(t) = N \; \forall t$, β is the transmission (infection) rate and γ is the recovery rate.

Suppose we have Gamma distributions on the parameters and $p(\beta, \gamma) = p(\beta)p(\gamma)$. Interest lies in the distribution of $\mathcal{R}_0 = \frac{\beta N}{\gamma}$.

Dynamic model example – Priors



(a)



(b)

Pool-then-induce

- Pool:

$$\pi(\beta) = Gamma(\theta_1^*, k_1^*)$$
$$\pi(\gamma) = Gamma(\theta_2^*, k_2^*)$$

  where $\theta^* = \sum_{i=0}^{K} \alpha_i a_i$. Then

- Induce:

$$\pi_1(\mathcal{R}_0) \propto {\mathcal{R}_0}^{k_1-1}(\theta_2^* \mathcal{R}_0 + N\theta_1^*)^{-(k_1^*+k_2^*)}$$

- Nice!

Induce-then-pool

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

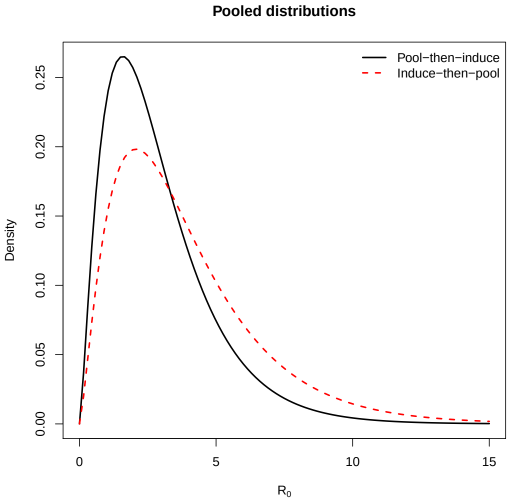
$$\pi_i(\mathcal{R}_0) \propto {\mathcal{R}_0}^{k_1 - 1}(\theta_2 \mathcal{R}_0 + N\theta_1)^{-(k_1+k_2)}$$

  then
- Pool:

$$\pi_2(\mathcal{R}_0) \propto \prod_{i=0}^{K} \pi_i(\mathcal{R}_0)^{\alpha_i}$$

- Ugly!

Dynamic model example (cont.)



Thank you!

- Thank you very much for your attention!
- The authors would like to thank Professor Adrian Raftery (University of Washington) 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

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

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.

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.