Combining spatio-temporal model predictions through logarithmic pooling

Luiz Max de Carvalho[a], Alvaro Faria[b]

a – School of Applied Mathematics, Getulio Vargas Foundation (FGV), Brazil
b – School of Mathematics & Statistics, The Open University, United Kingdom

January 28, 2021

Abstract

In many applied contexts, one has an ensemble of models which generate predictions, that must in turn be aggregated in order to form a consensus prediction. In particular, here we are concerned with *probabilistic predictions*, i.e., predictions in the form of probability statements or probability distributions. Logarithmic pooling is method for combining distributions that presents many attractive properties, but whose application in practice is hindered by lack of tractability of the normalising constant. The goal of the present project is to develop techniques to combine posterior predictive distributions from spatio-temporal models using output from Markov chain Monte Carlo (MCMC).

1 Background

Mathematical and statistical models are routinely used in decision-making, from weather prediction to epidemic forecasting. Since different models usually capture different features of the data, it is usual to combine multiple models in ensemble. Model ensembles usually provide more robust predictions and have found use in Physics (Leutbecher and Palmer, 2008), Environmetrics (Roy Bhowmik and Durai, 2008), Ecology (Araújo and New, 2007) and Epidemiology (Rosenfeld et al., 2012).

Suppose we have K models, $\boldsymbol{\mathcal{M}} = \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$ we would like to combine the predictions of. In particular, each model $\mathcal{M}_i$ is associated with a posterior predictive distribution:

$$f_i(\tilde{\boldsymbol{x}} \mid \boldsymbol{x}) = \int_{\boldsymbol{\Theta}} L_i(\tilde{\boldsymbol{x}} \mid \theta) L_i(\boldsymbol{x} \mid \theta) \pi_i(\theta) \, d\theta. \tag{1}$$

The goal is to combine the f_i in order to obtain a pooled predictive distribution p that reflects aspects of all the models considered and provides better predictions than any of the models in isolation.

2 Proper scoring rules

Assessing the quality of model predictions necessitates a principled way to *score* predictions in relation with the realised values. The so-called *proper scoring rules* provide a theoretically sound device to evaluate model predictions. The presentation here will follow Gneiting and Raftery (2007) very closely.

Let $\mathcal{P}$ be a convex class of probability measures on $(\Omega, \mathcal{A})$. We call $P \in \mathcal{P}$ a probabilistic forecast.

Definition 1. *We say $S : \Omega \to [-\infty, \infty]$ is a **scoring rule** if it is measurable and $S(P, \cdot)$ is P-quasi-integrable for all $P \in \mathcal{P}$.*

An interpretation of a scoring rule is that if a state $\omega \in \Omega$ is realised and the prediction was P, the reward to the forecaster is $S(P, \omega)$. The expected score under $Q \in \mathcal{P}$ if the forecast is P is

$$S(P, Q) := \int_{\Omega} S(P, \omega) dQ(\omega).$$

Definition 2. *We say S is **proper** if $S(Q, Q) \geq S(P, Q)$ for all $P, Q \in \mathcal{P}$. In addition, we say S is **strictly proper** if equality is achieved only for $P = Q$.*

Next, comes a crucial definition for this short note.

Definition 3. *If S is a (strictly) proper scoring rule and $c \geq 1$ is a constant and h is a $\mathcal{P}$-integrable function, then (Gneiting and Raftery, 2007, Eq. 2)*

$$S^*(P, \omega) = cS(P, \omega) + h(\omega) \tag{2}$$

*is also a (strictly) proper scoring rule and we say S^* and S are **strongly equivalent** if $c = 1$.*

Let μ be a σ-finite measure on $(\Omega, \mathcal{A})$. For $\alpha > 1$ define $\mathcal{L}_\alpha$ be the space of probability measures on $(\Omega, \mathcal{A})$ such that $\nu \ll \mu$ and $p(\omega) = \frac{d\nu}{d\mu}(\omega)$ and

$$||p||_\alpha = \left(\int_\Omega p(\omega)^\alpha d\mu(\omega) \right)^\alpha < \infty.$$

Of particular interest here is the logarithmic score

$$\mathrm{LogS}(p, \omega) = \log p(\omega), \tag{3}$$

also called the predictive deviance or the ignorance score. In the context of *local* scoring rules, that is, scoring rules $S(p, \omega)$ that depend on p only through the realised value ω, Bernardo (1979) showed that every proper local scoring rule is *equivalent* to the logarithmic score.

3 Logarithmic pooling

A central question when predictions in the form of probability distributions are available is how to combine them into a single pooled probability distribution. This is a topic of general interest, both in the statistical (West, 1984; Genest et al., 1986; Genest and Zidek, 1986) and decision theory literatures (Genest et al., 1984; French, 1985; Guardoni, 2002). A common situation of interest is combining predictive probability distributions about a quantity of interest $\theta \in \mathbf{\Theta} \subseteq \mathbb{R}^p$. Studying opinion pooling operators may give important insights on consensus belief formation and group decision making (West, 1984; Genest and Zidek, 1986; Guardoni, 2002). Among the various opinion pooling operators proposed in the literature, **logarithmic pooling** has enjoyed much popularity, mainly due to its many desirable properties such as relative propensity consistency (RPC), log-concavity (Carvalho et al., 2019) and external Bayesianity (EB) (Genest et al., 1986). 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).

First let us define the logarithmic pooling (LP) operator. Let $\mathbf{F}_\theta := \{f_1(\theta), f_2(\theta), \ldots, f_K(\theta)\}$ be a set of (densities of) prior distributions representing the opinions of $K + 1$ experts and let $\boldsymbol{\alpha} := \{\alpha_1, \alpha_2, \ldots, \alpha_K\}$ be the vector of weights, such that $\alpha_i > 0 \ \forall i$ and $\sum_{i=1}^K \alpha_i = 1$. The log-pooled prior is

$$\mathcal{LP}(\mathbf{F}_\theta, \boldsymbol{\alpha}) := \pi(\theta \mid \boldsymbol{\alpha}) = t(\boldsymbol{\alpha}) \prod_{i=1}^K f_i(\theta)^{\alpha_i}, \tag{4}$$

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

Logarithmic pooling can be shown to yield proper probability distributions for a finite number of densities (Genest et al., 1986, pg.489). It can also be shown that logarithmic pooling is the only aggregation method to universally preserve log-concavity, for any configuration of the

weights ($\boldsymbol{\alpha}$). This universality result is important because it holds for any set of log-concave distributions, $\mathbf{F}_\theta$. In contrast, a linear pool of K Gaussian distributions with common mean, for example, would produce a log-concave pooled distribution for any $\boldsymbol{\alpha}$, but this would fail if the means were different.

Consider the predictives described in (1) and their unnormalised equivalents, $\boldsymbol{F}^\star = \{f_1^\star, f_2^\star, \ldots, f_K^\star\}$, $f_i(\tilde{x} \mid x) = f_i^\star(\tilde{x} \mid x)/Z_i$. Applying the log-score in (3) to the logarithmic pooling of $\boldsymbol{F}^\star$ we have

$$\mathrm{LogS}^\star(\pi, \tilde{x}) = \sum_{i=1}^{K} \alpha_i \log f_i^\star(\tilde{x} \mid x), \tag{5}$$

which can easily be shown to be a (strictly) proper scoring rule. This is convenient because it allows the computation of a proper scoring rule without having to evaluate any normalising constants.

Open Problems

Using logarithmic pooling to combine predictive distributions is thus an exciting area of research, with many practical and theoretical lines of inquiry. Here, we describe two such lines of research, that would lead to direct practical gain, as well as necessitate novel theoretical approaches.

3.1 Sampling from the pooled predictive when only samples from previously obtained samples

An important setting is a situation where one has fitted K models to a set of data and thus obtained M samples for each of the K posterior predictive distributions. These samples are usually obtained *via* computationally-intensive MCMC techniques. One is only able to evaluate the unnormalised posterior densities $f_i^\star(\tilde{x} \mid x)$ and the Z_i are unknown. The challenge is then to produce samples from

$$\pi(\tilde{x} \mid x, \boldsymbol{\alpha}) = t(\boldsymbol{\alpha}) \prod_{i=1}^{K} f_i(\tilde{x} \mid x)^{\alpha_i}, \tag{6}$$

from a $M \times K$ matrix of samples

$$\tilde{\boldsymbol{X}} = \tilde{X}_1^{(1)}, \ldots, \tilde{X}_1^{(M)}$$
$$\tilde{X}_2^{(1)}, \ldots, \tilde{X}_2^{(M)}$$
$$\vdots$$
$$\tilde{X}_K^{(1)}, \ldots, \tilde{X}_K^{(M)}$$

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

While sampling from (6) directly is technically possible in certain situations, the challenge here is to produce samples from the pooled predictive by re-using the samples already available; posterior computation is laborious and ideally one would want to construct the pooled predictive with minimal computational overhead.

3.2 Calibrating the weights for optimal prediction

Another important research avenue, related to the above, is to combine the predictive distributions in an optimal way such that the combined predictions are more robust than any of those produced by the models individually. This involves optimising expressions of the form (5) to find

the vector $\boldsymbol{\alpha}^*$ that maximises predictive power according to the proper scoring rule of interest. This finds direct application in model ensembles in Weather and Epidemiology forecasting; employing the techniques would lead to optimal predictive distributions that are mathematically coherent, respect (strong) log-concavity constraints and provide predictions robust to model idiosyncrasies.

References

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

Araújo, M. B. and New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution*, 22(1):42–47.

Bernardo, J. M. (1979). Expected information as expected utility. *the Annals of Statistics*, pages 686–690.

Carvalho, L. M., Villela, D. V., Coelho, F. C., and Bastos, L. S. (2019). On the choice of weights for the logarithmic pooling of probability distributions. *In preparation*.

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.

French, S. (1985). Group consensus probability distributions: A critical survey in bayesian statistics. *Bayesian statistics*, 2.

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.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Guardoni, G. L. (2002). On irrelevance of alternatives and opinion pooling. *Brazilian Journal of Probability and Statistics*, pages 87–98.

Leutbecher, M. and Palmer, T. N. (2008). Ensemble forecasting. *Journal of computational physics*, 227(7):3515–3539.

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

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.

Rosenfeld, R., Grefenstette, J., and Burke, D. (2012). A proposal for standardized evaluation of epidemiological models.

Roy Bhowmik, S. and Durai, V. (2008). Multi-model ensemble forecasting of rainfall over indian monsoon region. *Atmosfera*, 21(3):225–239.

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.

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