Statistics and Probability Letters 119 (2016) 170–180

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bayesian aggregation of two forecasts in the partial information framework

Philip Ernst [a,*], Robin Pemantle [b], Ville Satopää [c], Lyle Ungar [d]

[a] Department of Statistics, Rice University, Houston, TX 77005, USA

[b] Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA

[c] Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, PA 19104, USA

[d] Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form 19 July 2016
Accepted 19 July 2016
Available online 4 August 2016

MSC:
primary 62C10
secondary 60G15

Keywords:
Expert
Probability forecast
Gaussian process
Judgmental forecasting

ABSTRACT

We generalize the results of Satopää et al(in press, 2015) by showing how the Gaussian aggregator may be computed in a setting where parameter estimation is not required. We proceed to provide an explicit formula for a "one-shot" aggregation problem with two forecasters.

1. Introduction

Prediction polling is a form of polling that asks a group of people to predict a common quantity. These forecasts are often used to make important decisions in medicine, economics, government, etc. In many practical settings, it is not possible to determine ex-ante which of the forecasters is the most informed or accurate (and even if this could be done, a decision to follow a specific forecaster's advice may result in relevant information from other forecasters being ignored). A more prudent solution is to pool the forecasters' information into a single consensus. This requires aggregators which can incorporate different information structures amongst the forecasters. This task motivated the work of Satopää et al. (in press, 2015), which introduced the Gaussian partial information framework for forecast aggregation. Further methodological framework for estimating parameters in the Gaussian partial information model was developed in Satopää et al. (2015).

The purpose of this letter is to further generalize the results of Satopää et al. (in press, 2015) by showing how the Gaussian aggregator may be computed via a (Bayesian) approach in which parameter estimation is not required. Our main result is Theorem 3.1, which provides an explicit formula for the Gaussian aggregator in a "one-shot" (a setting in which a stream of forecasts is unavailable) aggregation problem with two forecasters.

* Corresponding author.
E-mail addresses: philip.ernst@rice.edu (P. Ernst), pemantle@math.upenn.edu (R. Pemantle), satopaa@wharton.upenn.edu (V. Satopää), ungar@cis.upenn.edu (L. Ungar).

http://dx.doi.org/10.1016/j.spl.2016.07.018
0167-7152/

In the remainder of the introduction we give a brief description of important challenges in event forecasting and in forecast aggregation. We proceed to summarize the partial information framework, the Gaussian partial information model, and our Bayesian approach. Section 2 recalls the relevant computations for the Gaussian model with fixed parameters. Section 3 computes the Bayesian aggregator and Section 4 utilizes hypothetical data to compare the aggregators.

1.1. *Event forecasting, loss functions, and calibration*

In event forecasting, an expert is asked for a series $\{p_n\}$ of probability forecasts for events $\{A_n\}$. The quantitative study of event forecasting dates back at least three decades (Dawid, 1982; Murphy and Winkler, 1987). Usually, the expert is scored by a loss function $L(p_n, \mathbf{1}_{A_n})$. The loss function L is assumed to be *proper*, meaning that p minimizes $\mathbb{E}L(\cdot, Y)$ when Y is a Bernoulli random variable with mean p. Thus a forecaster with subjective probability p minimizes expected loss by forecasting p. For a more complete discussion of probability forecasting and proper loss functions, one may consult (Hwang and Pemantle, 1997).

Probability forecasts can suffer from two kinds of errors: bias and imprecision. Bias occurs when the long run frequency of A_n for those $p_n \approx p$ is not equal to p. Imprecision occurs when p_n is typically not close to zero or one. Assuming a sufficiently long stream of forecasts, each forecast p_n may be replaced by the forecast $q(p_n)$ where $q(t)$ is the long run frequency of A_n given a forecast of t. The forecast is then said to be *calibrated*; (cf. Murphy and Winkler, 1987) in this work we always assume calibrated forecasts. Of course, there are settings in which a stream of forecasts may not be available. In such a setting it is impossible to assess bias. A reasonable protocol is to assume no bias and to encourage calibration via proper loss functions (see Ungar et al., 2012).

Unlike other aggregators, a distinct advantage of one-shot aggregators is their universality; they can employed when a stream of forecasts is unavailable. One-shot aggregators can also serve as an equally applicable yet a more principled alternative to common aggregators such as the average and median. The simplicity of the average and the median aggregators has long been attractive to practitioners. The key contribution of this letter is to encourage the use of more principled aggregation techniques by providing a partial information aggregator that, too, has a simple and closed-form expression.

1.2. *Forecast aggregation*

Various probability models have been implicitly or explicitly used for producing a synthesized forecast from a collection of expert forecasts. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and events $A \in \mathcal{F}$. As discussed in Satopää et al. (in press, 2015), an expert's forecast is considered to be calibrated if the forecast p for an event A is equal to $\mathbb{P}(A|\mathcal{F}')$ for some $\mathcal{F}' \subseteq \mathcal{F}$. The σ-field $\mathcal{F}'$ represents the information used to make the forecast; it need not be the full information available to the expert.

Some empirical work on forecast aggregation operates outside the above framework. For example, the *measurement error framework* assumes there is a true probability θ, interpreted as the forecast made by an "ideal" forecaster. The actual forecasters observe a transformation $\phi(\theta)$ together with independent mean zero idiosyncratic errors. This leads to relatively simple aggregation rules. For example, if ϕ is the identity, the forecasters are assumed to be reporting θ plus independent mean zero errors. The corresponding aggregator then simply averages the forecasts

$$g_{\text{ave}}(p_1, \dots, p_n) := \frac{1}{n}\sum_{k=1}^{n} p_k. \tag{1}$$

When the function ϕ is Φ^{-1} (the inverse normal CDF) this leads to *probit averaging*, defined by

$$g_{\text{probit}}(p_1, \dots, p_n) := \Phi\left(\frac{1}{n}\sum_{k=1}^{n} \Phi^{-1}(p_k)\right). \tag{2}$$

Such models, while very common in practice, lead both to uncalibrated forecasts and suboptimal performance. Some theoretical problems with these models are discussed by Hong and Page (2009); for example, such aggregators can never leave the convex hull of the individual expert forecasts, which is demonstrably sub-optimal in some cases (Parunak et al., 2013); see also Satopää et al. (2015, Section 2.3.2).

Letting $\mathcal{F}'' = \sigma(p_1, \dots, p_n)$, we define an aggregator as any random variable $\tilde{p} \in \mathcal{F}''$. Then, amongst all such aggregators, p'' (see (3)) is the one that minimizes the expectation of any proper loss function. It is also calibrated.

In the *partial information framework* for aggregation of calibrated forecasts proposed by Satopää et al. (in press, 2015), each forecaster i, $1 \le i \le N$ is assumed to have access to information $\mathcal{F}_i$. The aggregator only considers the forecasts $p_i := \mathbb{P}(A|\mathcal{F}_i)$. Theoretically, the best possible forecast with this information is the *revealed estimator*

$$p'' := \mathbb{P}(A|p_i : 1 \le i \le N). \tag{3}$$

It is clear that

$$p'' = g_{\text{rev}}(p_1, \dots, p_n)$$

for some function $g = g_{\text{rev}}$; however, it is not possible to explicitly compute g without making further assumptions about the model.