Supplementary Material for "Evaluating infectious disease forecasts with allocation scoring rules"

Aaron Gerding, Nicholas G. Reich, Benjamin Rogers, Evan L. Ray

December 21, 2023

Contents

- ☐ 2 actual references to supplement (re exponential quantiles and solving the allocation problem)
- ☐ the statement that "we computed the allocation score for the 14 day-ahead forecast" which needs some answer to "how?"
- ☐ use of "proper" a number of times

Abstract

We briefly address some technical and methodological points in the main text, referring to the forthcoming ... for more thorough discussion.

1 Introduction

We briefly address some technical and methodological points in the main text. We begin in section **??** by formalizing the concept of a *shortage* of resources and giving some key results about expected resource shortages under a distribution characterizing uncertainty about (future) levels of resource need. Resource shortages play a central role in the decision-making problems that give rise to the quantile loss and the allocation score, which we discuss in sections **??** and 4 respectively. Section 5 gives details on the numerical methods that we use to calculate allocation scores, including some special considerations for settings where forecasts are represented by a finite collection of predictive quantiles, such as the application to forecasts of hospitalizations due to COVID-19 in section 3 of the article.

2 Proper scoring rules

In decision theory, a loss function l is used to formalize a decision problem by assigning numerical value $l(x, y)$ to the *result* of taking an *action* x in preparation for an *outcome* y. A *scoring rule* S is a loss function for a decision problem where the action is a probabilistic forecast F of the outcome y (or the statement of F by a forecaster). We refer to the realized loss $S(F, y)$ as the *score* of F at y.

Decision theoretically, probabilistic forecasts are a unique kind of action in that they can be used to generate their own (simulated) outcome data, against which they can be scored using S. A probabilistic forecast F is thus committed to the "self-assessment" $\mathbb{E}_F[S(F, Y)] := \mathbb{E}[S(F, Y^F)]$, where $Y^F \sim F$ is the random variable defined by sampling from F, as well to an assessment $\mathbb{E}_F[S(G, Y)]$ of any alternative forecast G.

A natural consistency criterion for S is that it does not commit F to assessing any other forecast G as being better than F itself, that is, that

$$\mathbb{E}_F[S(F, Y)] \leq \mathbb{E}_F[S(G, Y)] \tag{1}$$

for any F, G. A scoring rule meeting this criterion is called *proper*. If S were improper, then from the perspective of a forecaster focussed (solely) on expected loss minimization (which we will call an ELM forecaster), the decision to state a forecast G other than the forecast F which they believe describes Y could be superior to the decision to state F. S is *strictly proper* when (1) is sharp, in which case the *only* optimal decision for an ELM forecaster is to state the forecast they believe to be true.

2.1 The allocation score is proper

Our primary decision theoretical procedure, outlined in section 2.2.1 of the main text, uses a decision problem with loss function $s(x, y)$ to define a scoring rule

$$S(F, y) := s(x^F, y) \tag{2}$$

where $x^F := \text{argmin}_x \mathbb{E}_F[s(x, Y)]$ is the Bayes act for F with respect to s. Such scoring rules, which we call *Bayes scoring rules*, are proper by construction since

$$\begin{aligned}
\mathbb{E}_F[S(F, Y)] &= \mathbb{E}_F[s(x^F, Y)] \\
&= \min_x \mathbb{E}_F[s(x, Y)] \quad \text{(by definition of } x^F) && (3) \\
&\leq \mathbb{E}_F[s(x^G, Y)] && (4) \\
&= \mathbb{E}_F[S(G, Y)].
\end{aligned}$$

The allocation scoring rule is Bayes and therefore proper.

We note that in the probabilistic forecasting literature (see e.g., Gneiting [2011], Theorem 3) what we have termed Bayes scoring rules typically appear via (2) where x^F is some given functional of F which can be shown to be *elicitable*, that is, to be the Bayes act for some loss function s. Such a loss function is said to be a *consistent loss (or scoring) function* for the functional $F \mapsto x^F$, and many important recent results in the literature (e.g., Fissler and Ziegel [2016]) address whether there *exists* any loss function for x^F which is consistent. Our orientation is different from this insofar as we *begin* by specifying a decision problem and a loss function of subject matter relevance and use the Bayes act only as a bridge to a proper scoring rule. Consistency is never in doubt.

3 Expected shortages

A key feature of loss functions used in the decision theoretic definition of quantiles and related scoring rules such as the CRPS and the WIS (see ...), as well as the allocation loss function presented in this work, is the presence of a *shortage*: the amount $\max\{0, y - x\}$ by which a random resource demand y exceeds a supply decision variable x, which, for convenience, we write as $(y - x)_+$. In particular, a quantile at probability level α of a distribution F on $\mathbb{R}^1$ (which we assume to have a well-defined density $f(x)$) is a Bayes act for the loss function

$$l(x, y) = Cx + L(y - x)_+$$

where $\alpha = 1 - C/L$ and C and L can be interpreted as the cost per unit of a resource (such as medicine) and the loss incurred when a unit of demand (such as illness) cannot be met due to the shortage $(y-x)_+$. This follows because a Bayes act, as a minimizer of $\mathbb{E}_F[l(x,Y)]$, must also be a vanishing point of the derivative

$$\begin{aligned}\frac{d}{dx}\,\mathbb{E}_F\left[l(x,Y)\right] &= \mathbb{E}_F\left[\frac{d}{dx}l(x,Y)\right]\\ &= C + L\,\mathbb{E}_F\left[\frac{d}{dx}(Y-x)_+\right]\\ &= C - L\,\mathbb{E}_F\left[\mathbf{1}\{Y>x\}\right]\\ &= C + L(F(x)-1),\end{aligned}\tag{5}$$

so that $1-C/L = F(x)$. The formula $\frac{d}{dx}\,\mathbb{E}_F\left[(Y-x)_+\right] = F(x)-1$ for the derivative of the shortage which we use below in deriving the Bayes act for the allocation loss, can be seen as following, as indicated above in (5), from a probability theoretic definition of the expectation operator on piecewise defined functions, or as application of the "Leibniz Rule"

$$\begin{aligned}\frac{d}{dx}\,\mathbb{E}_F[(Y-x)_+] &= \frac{d}{dx}\int_x^\infty (y-x)f_Y(y)dy\\ &= \int_x^\infty \frac{d}{dx}(y-x)f_Y(y)dy - (x-x)f_Y(x) = -\int_x^\infty f_Y(y)dy = F(x)-1.\end{aligned}\tag{6}$$

(Note that more care is required when F does not have a density.)

4 Allocation Bayes acts as vectors of marginal quantiles.

Here we show that the Bayes act $x^{F,K} = (x_1^{F,K}, \ldots, x_N^{F,K})$ for a forecast F, corresponding to the allocation problem (AP) ((3) in section 2.2.2)

$$\underset{0\le x}{\text{minimize}}\ \mathbb{E}_F[s_A(x,Y)] = \sum_{i=1}^N L\cdot\mathbb{E}_{F_i}[(Y_i-x_i)_+] \text{ subject to } \sum_{i=1}^N x_i = K,\tag{7}$$

can be represented as a vector of quantiles for the marginal forecast distributions F_i at a single probability level $\tau^{F,K}$, that is, $x_i^{F,K} = q_{F_i,\tau^{F,K}}$. An immediate consequence used in the examples in Section 2.1 in the main text is that if $F_i = \text{Exp}(1/\sigma_i)$ for all i, then the Bayes act is proportional to $(\sigma_1,\ldots,\sigma_N)$, since $q_{\text{Exp}(1/\sigma),\tau} = -\sigma\log(1-\tau)$.

In order for $x^\star \in \mathbb{R}_+^N$ to solve the AP it must be true that reallocatating $\delta > 0$ of the $x_i^\star$ units of resource allocated to location i to location j will increase the expected shortage in location i by at least as much as it decreases the expected shortage in location j. Letting $\delta \searrow 0$, this implies from (6) that

$$\begin{aligned}1-F_i(x_i^\star) &= -\frac{d}{dx_i}\mathbb{E}_{F_i}[(Y_i-x_i^\star)_+]\\ &= \lim_{\delta\searrow 0}\frac{1}{\delta}\left\{\mathbb{E}_{F_i}[(Y_i-(x_i^\star-\delta))_+] - \mathbb{E}_{F_i}[(Y_i-x_i^\star)_+]\right\} \text{ (increase in } i)\\ &\ge \lim_{\delta\searrow 0}\frac{1}{\delta}\left\{\mathbb{E}_{F_j}[(Y_j-x_j^\star)_+] - \mathbb{E}_{F_j}[(Y_j-(x_j^\star+\delta))_+]\right\} \text{ (decrease in } j)\\ &= -\frac{d}{dx_j}\mathbb{E}_{F_j}[(Y_j-x_j^\star)_+] = 1-F_j(x_j^\star)\end{aligned}\tag{8}$$

Note that negative derivatives appear because our optimality condition addresses how a *decrease* in resources will *increase* the expected shortage in i and vice versa in j. Since (8) holds with i and j reversed, a number λ (a *Lagrange multiplier*) exists such that $L(1-F_k(x_k^\star)) = \lambda$ for all $k \in 1,\ldots,N$. (We scale by L to facilitate possible future interpretations of λ in terms of the partial derivatives of

$\mathbb{E}_F[s_A(x, Y)]$.) That is, $x_k^\star$ is a quantile q_{τ,F_k} for $\tau = 1 - \lambda/L$. The value of τ is then determined by the constraint equation

$$\sum_{i=1}^{N} q_{\tau,F_i} = K. \tag{9}$$

It is important to note that τ depends on F and K and is *not* a fixed parameter of the allocation scoring rule.

5 Numerical computation of allocation Bayes acts

To compute an allocation score $S_A(F, y; K) := s_A(x^{F,K}, y)$, we require numerical values for a Bayes act solving the AP (7). Assuming we have reliable means of calculating quantiles q_{α,F_i} of the marginal forecasts F_i, such values are given by $q_{\tau^\star,F_i}$ where $\tau^\star$ solves the equation (9). Usually, however, this equation is not analytically tractable and we must resort to a numerical method for finding an approximation $\tilde{\tau}$ of $\tau^\star$. This task is made simpler by $\sum_{i=1}^N q_{\tau,F_i}$ being an increasing function of τ, and, in principle, a bisection method that evaluates the sign of $\sum_{i=1}^N q_{\tau,F_i} - K$ at each midpoint of the search interval, narrowing it accordingly, will suffice. We have implemented such a procedure along with the resulting score computations in the R package `alloscore` which provided all allocation score values used in the analysis of section 3 in the main text.

Awkward subtleties can arise though when the forecast densities f_i vanish or are very small, in which case quantiles are non-unique or highly variable at a probability level, leading to ambiguity in how to evaluate $\sum_{i=1}^N q_{\tau,F_i}$. And if point masses are present in any of the F_i, (9) will fail to have a unique solution for some discrete set of constraint levels K. Potentially complex conventions must be adopted for both detecting such levels and enforcing consistency in score calculations near them. In both situations, it appears to us that conditions guaranteeing that numerical instabilities and artifacts do not threaten the validity of inferences regarding forecaster performance may be highly dependent on N as well as forecast ranges and regularity. We have established and coded conditions that through extensive experimentation seem appropriate for... but we leave a more rigorous approximation error analysis for later work.

6 Computing allocations from finite quantile forecast representations

In section 3 of the manuscript, we used the allocation score to evaluate forecasts of COVID-19 hospitalizations that have been submitted to the COVID-19 Forecast Hub. These forecasts are submitted to the Hub using a set of 23 quantiles of the forecast distribution at the 23 probability levels in the set $\mathcal{T} = \{0.01, 0.025, 0.05, 0.1, 0.15, \ldots, 0.9, 0.95, 0.975, 0.99\}$, which specify a predictive median and the endpoints of central $(1-\alpha) \times 100\%$ prediction intervals at levels $\alpha = 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. For a given week and target date, we use $q_{i,k}$ to denote the submitted quantiles for location i and probability level $\tau_k \in \mathcal{T}$, $k = 1, \ldots, 23$.

In the event that there is some $k \in \{1, \ldots, 23\}$ for which $\sum_i q_{i,k} = K$, i.e., the provided predictive quantiles at level τ_k sum across locations to the resource constraint K, the solution to the allocation problem is given by those quantiles. However, generally this will not be the case; the optimal allocation will typically be at some probability level $\tau^\star \notin \mathcal{T}$.

To address this situation and support the numerical allocation algorithm outlined in section 5, we need a mechanism to approximate the full cumulative distribution functions F_i, $i = 1, \ldots, N$ based on the provided quantiles. We have developed functionality for this purpose in the `distfromq` package for R (cite). This functionality represents a distribution as a mixture of discrete and continuous parts, and it works in two steps:

1. Identify a discrete component of the distribution consisting of zero or more point masses, and create an adjusted set of predictive quantiles for the continuous part of the distribution by subtracting the point mass probabilities and rescaling.

2. For the continuous part of the distribution, different approaches are used on the interior and exterior of the provided quantiles:
   (a) On the interior, a monotonic cubic spline interpolates the adjusted quantiles representing the continuous part of the distribution.
   (b) A location-scale parametric family is used to extrapolate beyond the provided quantiles. The location and scale parameters are estimated separately for the lower and upper tails so as to obtain a tail distribution that matches the two most extreme quantiles in each tail.

The resulting distributional estimate exactly matches all of the predictive quantiles provided by the forecaster. We use the cumulative distribution function resulting from this procedure as an input to the allocation score algorithm.

We refer the reader to the `distfromq` documentation for further detail.

7 Propriety of parametric approximation

Bayes scoring rules are invariant under reparametrization of the action space.

Let $S(F, y)$ be a Bayes scoring rule derived from an action set $\mathcal{X}$ and loss $s(x, y)$. Let $\mathcal{W}$ be another set and $M : \mathcal{W} \to \mathcal{X}$ a map. M defines a loss $s_M(\theta, y) := s(M(\theta), y)$ on $\mathcal{W}$. This in turn defines the Bayes scoring rule

$$S_M(F, y) := s_M(\theta^F, y) = s(M(\theta^F), y). \tag{10}$$

If M is surjective onto the set of Bayes acts $\{x \in \mathcal{X} \mid x = x^F \text{ for some } F \in \mathcal{P}\}$, then the scores coincide since

$$\begin{aligned} S_M(F, y) = s_M(\theta^F, y) &= \min_\theta \{s_M(\theta, F)\} &(11)\\ &= \min_\theta \{s(M(\theta), F)\} &(12)\\ &= \min_x \{s(x, F)\} = s(x^F, F) = S(F, y). &(13) \end{aligned}$$

(Note that this does not imply that $M(\theta^F) = x^F$ since x^F might not be a unique Bayes act.)

Identifying $G_\theta \in \mathcal{P}_{\text{dfq}}$ with its quantile parameters θ, the map $M(\theta) := x^\theta$ is surjective onto the action set of allocations under any constraint K since they can be specified as quantiles for distributions in $\mathcal{P}_{\text{dfq}}$. Therefore the `distfromq` "reparametrized" allocation scoring rule $S_{A,\text{dfq}}(F, y) := s_{\text{dfq}}(\theta^F, y) = s_A(x^{\theta^F}, y)$ coincides with the AS $S_A(F, y)$ itself.

Unfortunately,

For convenience we write $S(F, G) := \mathbb{E}_G[S(F, Y)]$. Let

$$\begin{aligned} q_{23}(F) &:= (q_{.01,F}, \ldots, q_{.99,F}) \\ G^q(F) &:= G_{q_{23}(F)} \in \mathcal{P}_{\text{dfq}} \\ G^\star(F) &:= \operatorname*{argmin}_{G \in \mathcal{P}_{\text{dfq}}} S_A(G, F) \end{aligned}$$

We have the "right" approximate allocation score

$$S_R(F, y) := S_A(G^\star(F), y)$$

which is proper because

$$\begin{aligned} S_R(F, F) &= S_A(G^\star(F), F) \\ &\le S_A(G^\star(F), H) \quad \text{(by definition of } G^\star) \\ &= S_R(F, H). \end{aligned}$$

This seems to almost be what is called "properization" in Brehmer and Gneiting [2020]. We also have a "wrong" approximation

$$\tilde{S}_R(F, y) := S_A(G^q(F), y)$$

which is improper because

$$\begin{aligned}\tilde{S}_R(F, G^q(F)) &= S_A(G^q(F), G^q(F)) \\ &< S_A(G^q(F), F) \quad \text{(because } S_A \text{ is proper)} \\ &= \tilde{S}_R(F, F).\end{aligned}$$

References

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

Tobias Fissler and Johanna F. Ziegel. Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4):1680 – 1707, 2016.

Jonas R Brehmer and Tilmann Gneiting. Properization: constructing proper scoring rules via bayes acts. *Annals of the Institute of Statistical Mathematics*, 72:659–673, 2020.