

Supplementary Material for “Evaluating infectious disease forecasts with allocation scoring rules”

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

December 20, 2023

Contents

1	Introduction	1
2	Proper scoring rules	1
2.1	The allocation score is proper	2
3	Expected shortages	2
4	Allocation Bayes acts as vectors of marginal quantiles.	3
5	Numerical computation of allocation Bayes acts	4
6	Computing allocations from finite quantile forecast representations	5
7	Propriety of parametric approximation	5
<input type="checkbox"/>	2 actual references to supplement (re exponential quantiles and solving the allocation problem)	
<input type="checkbox"/>	the statement that “we computed the allocation score for the 14 day-ahead forecast” which needs some answer to “how?”	
<input type="checkbox"/>	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

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

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 ?? of the main text, uses a decision problem with loss function $s(x, y)$ to define a scoring rule

$$S(F, y) := s(x^F, y) \quad (2)$$

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

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

$$\begin{aligned} &\leq \mathbb{E}_F[s(x^G, Y)] \\ &= \mathbb{E}_F[S(G, Y)]. \end{aligned} \quad (4)$$

The allocation scoring rule is Bayes and therefore proper.

We note that in the probabilistic forecasting literature (see e.g., ?, 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., ?) 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

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

so that $1 - C/L = F(x)$. The formula $\frac{d}{dx} \mathbb{E}_F[(Y - x)_+] = F(x) - 1$ 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 $x^{F,K} = (x_1^{F,K}, \dots, x_N^{F,K})$ for a forecast F , corresponding to the allocation problem (AP) ((??) in section ??)

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

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 ?? in the main text is that if $F_i = \text{Exp}(1/\sigma_i)$ for all i , then the Bayes act is proportional to $(\sigma_1, \dots, \sigma_N)$, since $q_{\text{Exp}(1/\sigma), \tau} = -\sigma \log(1 - \tau)$.

In order for $x^* \in \mathbb{R}_+^N$ to solve the AP it must be true that reallocatating $\delta > 0$ of the x_i^* 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

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

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 $L(1 - F_k(x_k^*)) = \lambda$ for all $k \in 1, \dots, N$. (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^* is a quantile q_{τ, F_k} for $\tau = 1 - \lambda/L$. The value of τ is then determined by the constraint equation

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

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

In practice, (??) is rarely amenable to closed form solution. Using the Bayes act in explicit computations therefore requires a method of generating an approximation $\tilde{\tau}$ to a solution τ^* of (??). One would also strive to have some control over a robust measure of the approximation error of the random shortages which we aim to compute using the Bayes act. One possible measure is the sum of absolute coordinate errors

$$\text{SAE}(Y, \tilde{\tau}) = \sum_{i=1}^N \left| (Y_i - q_{\tilde{\tau}, F_i})_+ - (Y_i - x_i^{F, K})_+ \right| \quad (10)$$

where $q_{\tilde{\tau}, F}$ solves the linear program $\sum_{x \in \mathbf{F}^{-1}(\tilde{\tau})} x_i = K$ approximating the true Bayes act defining linear program $\sum_{x \in \mathbf{F}^{-1}(\tau^*)} x_i = K$. In this section we outline such a method we have implemented in the R package `alloscore`.

Suppose we have established that $\tau^* = F_T(K)$ lies in the interval $I_1 = [\tau_L, \tau_U]$ with $\tau_L < \tau_U$, that is, $K \in [q_{F_T, \tau_L}^-, q_{F_T, \tau_U}^+]$. From section ??, we know that the set $TQ_F(\tau_L) \cup TQ_F(\tau_U) \subset [q_{F_T, \tau_L}^-, q_{F_T, \tau_U}^+]$ is arranged in exactly one of the following ways (where e.g. refers to figure ??):

- (••) $TQ_F(\tau_L) \cap TQ_F(\tau_U) = \emptyset$ (and $q_{F_T, \tau_L}^+ < q_{F_T, \tau_U}^-$, e.g., $\tau_L = \alpha_1, \tau_U = \alpha_5$)
- (•) $TQ_F(\tau_L) = \{K\} = TQ_F(\tau_U)$ (a point mass at K , e.g., $\alpha_3 \leq \tau_L \leq \tau_U \leq \alpha_5$)
- (•–) $TQ_F(\tau_L) = \{q_{F_T, \tau_U}^-\} \subsetneq TQ_F(\tau_U)$ (a point mass at q_{F_T, τ_U}^- , e.g., $\tau_L = \alpha_3, \tau_U = \alpha_5$)
- (–•) $TQ_F(\tau_L) \supsetneq \{q_{F_T, \tau_L}^+\} = TQ_F(\tau_U)$ (a point mass at q_{F_T, τ_U}^+ , e.g., e.g., $\tau_L = \alpha_5, \tau_U = \alpha_4$).

In the case (•), we can immediately take $\tau^* = \tau_U$ as the probability level representing the allocation Bayes act. In the cases (•–), and (–•), which imply the presence of a point mass in one or more of the component forecasts adjacent to a region of zero density in all components, we can take $\tau^* = \tau_U$ or τ_L , respectively, as the representing probability level. Having found our τ^* we then arrive at a Bayes act $x^{F, K}$ by solving the linear program $\sum_{x \in \mathbf{F}^{-1}(\tau^*)} x_i = K$.

In the remaining and typical case of (••), further search is generally necessary. We can proceed by evaluating TQ_F at $\tau_M = \frac{1}{2}(\tau_L + \tau_U)$ and replacing I_1 with one of its halves

$$I_2 = \begin{cases} [\tau_L, \tau_M] & \text{if } K < q_{F_T, \tau_M}^- \\ [\tau_M, \tau_U] & \text{if } K \geq q_{F_T, \tau_M}^- \end{cases} \quad (11)$$

which also contains $\tau^* = F_T(K)$. This follows from the definition $q_{F_T, \tau_M}^- := \min\{x \mid F_T(x) \geq \tau_M\}$, according to which

$$\begin{cases} K < q_{F_T, \tau_M}^- \text{ implies } \tau^* = F_T(K) < \tau_M \\ K \geq q_{F_T, \tau_M}^- \text{ implies } \tau^* = F_T(K) \geq \tau_M. \end{cases} \quad (12)$$

Note that $K \leq q_{F_T, \tau_M}^-$ would not imply $F_T(K) \leq \tau_M$ due to the possibility of a point mass at K .

Iterating this process we obtain a sequence $\{I_k\}, k = 1, 2, \dots$ of intervals of widths $|I_k| = 2^{1-k} |I_1|$ which either terminates at one of the scenarios (•), (•–), or (–•), or provides infinite sequences $\{\tau_{L,k}\}$ and $\{\tau_{U,k}\}$ converging to τ^* from below and above. Such a sequence provides the basis of a “bisection” algorithm for finding the “root” τ^* of the set condition $0 \in TQ_F(\tau) - K$.

In the generic case of an infinite $\{I_k\}$, we need to define practical stopping criteria for the possible limit behaviours of $F_T(K \pm \varepsilon)$ as $\varepsilon \searrow 0$ which are exemplified in figure ?? at $0, x_1, x_2, x_3, (x_3 + x_4)/2$ and x_5 .

And after having obtained sufficiently a sufficient narrow terminal $I_K = [\tau_{L,K}, \tau_{U,K}]$ we need a protocol for how to select a solution to the linear program $\sum_{x \in \mathbf{F}^{-1}(I_K)} x_i = K$ as an approximate Bayes act.

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 $\mathcal{T} = \{0.01, 0.025, 0.05, 0.1, 0.15, \dots, 0.9, 0.95, 0.975, 0.99\}$, 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 $\tau_k \in \mathcal{T}$, $k = 1, \dots, 23$.

In the event that there is some $k \in \{1, \dots, 23\}$ 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 $\tau^* \notin \mathcal{T}$.

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 , $i = 1, \dots, N$ 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

In practice, open forecasting exercises are generally not able to collect an exact representation of the forecast distribution F other than in simple settings such as for a categorical variable with a relatively small number of categories. In settings where the outcome being forecasted is a continuous quantity (such as rates of influenza-like illness among outpatient doctor visits) or a count (such as influenza hospitalizations), forecasting exercises have therefore resorted to collecting summaries of a forecast distribution such as bin probabilities or predictive quantiles. This raises the question of whether use of the allocation score still constitutes a proper evaluation procedure if the forecast distribution F is not itself directly recorded. The answer we provide in this section is that it is, so long as the fact that the allocation score will be used for forecast evaluation and the method that will be used to obtain a distributional estimate of F from the provided forecast representation are communicated to participating forecasters prospectively.

We consider a setting where a forecasting exercise (such as a forecast hub) pre-specifies that forecasts will be represented using a parametric family of forecast distributions $G_\theta(y)$, and the task of the forecaster is to select a particular parameter value θ . We use \mathcal{P} to denote the collection of all distributions

G_θ in the given parametric family. For instance, it has recently been proposed that mixture distributions could be used to represent forecast distributions ?. Additionally, we note that the functionality in `distfromq` can be viewed as specifying a parametric family \mathcal{P}_{dfq} where the parameters θ of G_θ are its quantiles at pre-specified probability levels, and where the shape of any $G_\theta \in \mathcal{P}_{\text{dfq}}$ over the full range of its support is entirely controlled by these quantiles.

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

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

We have the “right” approximate allocation score

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

which is proper because

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

This seems to almost be what is called “properization” in ?. We also have a “wrong” approximation

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

which is improper because

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

As another alternative for practical forecasting exercises, a forecast hub could ask forecasters to directly provide the Bayes allocations associated with their forecasts for one or more specified resource constraints K . At the cost of increasing the number of quantities solicited by the forecast hub, this would have several advantages: it would prevent any artificial distortion of the forecast distributions, allow for direct calculation of scores, and narrow the gap between model outputs and public health end users. For this to be feasible, implementations of the allocation algorithm would have to be provided to participating forecasters in the computational languages being used for modeling.