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

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

December 12, 2023

Abstract

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

- ☑ From 2.2.1, why are Bayes act scoring rules proper?
- ☐ DGP as optimal for any decision problem, ref Diebold, Gunther, Tay p. 866; and if forecasts are ideal, then forecasts with better information always yield better decisions, ref Holzmann and Eulert, Corr 2.
- ☑ For 2.2.2, how to get quantile representation of Bayes act using Lagrange multiplier, ~~assuming smooth, never-zero densities well behaved at $x = 0$.~~ Work out exponential example. Refer to methods paper for general case.
- ☑ Derivation of quantile scoring rule with quantile as Bayes act for C/L problem, ~~assuming never-zero densities~~.
- ☐ algorithmic details
  - ☐ use of `distfromq` to get from quantiles to distribution functions
  - ☐ `alloscore`
  - ☐ implications for propriety. do quantiles elicited by `distfromq` $\hookrightarrow$ `alloscore` process align with "real" quantiles? the alloscore is proper if distribution functions F are handed to us; is it still proper given our algorithm situation?
- ☐ Descriptions of
  - ☐ CRPS as average quantile score across $C \in [0/L]$ decision problems
  - ☐ IS as average of two quantile scores with a prob-width penalty
  - ☐ WIS as average quantile score across 23 C/L problems.
- ☐ Derivation of case 2 in formula for Oracle adjustment

1 Introduction

We briefly address some technical and methodological points in the main text. We begin in section 2 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 3 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 Shortages

We use y to denote the demand or need for resources and x to denote the level of available resources. The resource shortage is the amount by which resource demand exceeds supply. For convenience, we write $u_+ = \max\{0, u\}$, i.e., "the positive part" of u. With this notation, the shortage is written as $(y-x)_+$. To regard shortage as a function of only one variable x or y, with the other being a parameter describing the dependence we can write $(y-x)_+ = \mathrm{sh}^y(x) = \mathrm{sh}_x(y)$. Note that $\mathrm{sh}^y(x)$ and $\mathrm{sh}_x(y)$ are both convex functions and "mirror" each other (see figure 1).

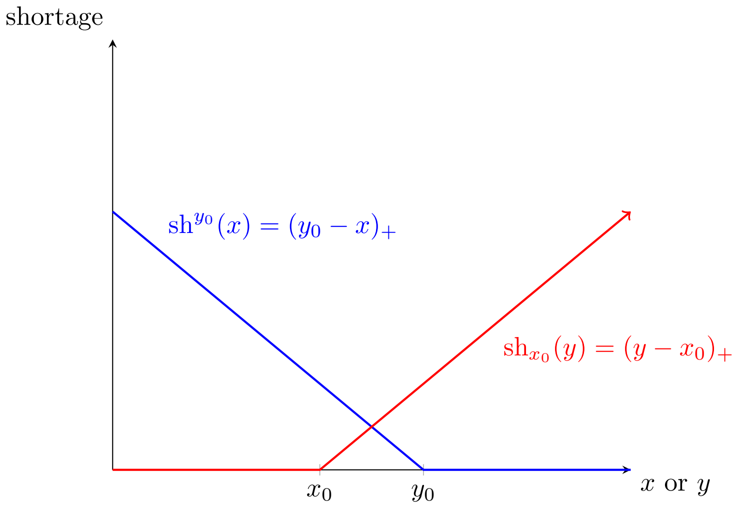

Figure 1: Shortage functions. $\mathrm{sh}^{y_0}(x)$, shown in blue, gives the resource shortage as a function of the level of available resources, x, for a fixed value of resource demand y_0. $\mathrm{sh}_{x_0}(y)$, shown in red, gives the resource shortage as a function of resource demand, y, for a fixed value of resource supply x_0.

Let Y be a random variable with distribution F representing the unknown level of resource demand. The random shortage $(Y-x)_+$ can be thought of as either a real-valued random variable $\mathrm{sh}_x(Y)$ for every x, or a function-valued random variable sh^Y whose value for any realization $Y = y$ is a convex function $\mathrm{sh}^y(x)$ of x. In the sections below, we will work with the *expected shortage*[1] $\mathbb{E}_F[(Y-x)_+] = \mathbb{E}_F[\mathrm{sh}^Y](x)$. Assuming that this expected value exists, which is the case as long as the distribution F is well behaved, we can see that $\mathbb{E}_F[\mathrm{sh}^Y](x)$ is also convex (and therefore continuous) in x by integrating the convexity inequality for $\mathrm{sh}^y(x)$ with respect to the probability measure $dF(y)$:

$$\begin{aligned}
\mathbb{E}_F[\mathrm{sh}^Y](\lambda x_1 + (1-\lambda)x_2) &= \int \mathrm{sh}^y(\lambda x_1 + (1-\lambda)x_2) dF(y) \\
&\leq \int \lambda\, \mathrm{sh}^y(x_1) + (1-\lambda)\, \mathrm{sh}^y(x_2) dF(y) \\
&= \lambda\, \mathbb{E}_F[\mathrm{sh}^Y](x_1) + (1-\lambda)\, \mathbb{E}_F[\mathrm{sh}^Y](x_2).
\end{aligned} \tag{1}$$

[1] A more natural sounding term for $(y-x)_+$ might have been *shortfall*. Unfortunately *expected shortfall* has long been used in finance to refer to quantities more closely related to the *conditional* expectation $\mathbb{E}_F[Y - x \mid Y - x \geq 0] = \mathbb{E}_F[(Y-x)_+]/\mathbb{P}_F\{Y \geq x\}$.

Convexity is also shown by directly exhibiting the the left and right derivatives of $\mathbb{E}_F[\mathrm{sh}^Y](x)$:

$$
\begin{aligned}
D_- \mathbb{E}_F[(Y-x)_+] &= \lim_{h \searrow 0} \frac{1}{h} \mathbb{E}_F[(Y-x)_+ - (Y-(x-h))_+] && (2)\\
&= \lim_{h \searrow 0} \frac{1}{h} \int_{[x-h,x]} (x-h-y) dF(y) - \lim_{h \searrow 0} \frac{1}{h} \int_{(x,\infty)} h dF(y) && (3)\\
&= \lim_{h \searrow 0} \frac{1}{h} \int_{[x-h,x]} -h dF(y) - 1 + F(x) && (4)\\
&= -(F(x) - F(x-)) - 1 + F(x) \quad \left(\text{where } F(x-) := \lim_{t \nearrow x} F(t)\right) && (5)\\
&= F(x-) - 1 && (6)\\
D_+ \mathbb{E}_F[(Y-x)_+] &= \lim_{h \searrow 0} \frac{1}{h} \mathbb{E}_F[(Y-(x+h))_+ - (Y-x)_+] && (7)\\
&= \lim_{h \searrow 0} \frac{1}{h} \int_{[x,x+h]} (x-y) dF(y) - \lim_{h \searrow 0} \frac{1}{h} \int_{(x+h,\infty)} h dF(y) && (8)\\
&= \lim_{h \searrow 0} \frac{1}{h} \int_{[x,x+h]} 0 dF(y) - 1 + F(x) && (9)\\
&= F(x) - 1 && (10)
\end{aligned}
$$

where in (4) and (9) we are able to replace the integrands with their values at x because they are bounded over the shrinking regions of integration $[x-h, x]$ and $[x, x+h]$. Convexity follows since $D_- \mathbb{E}_F[\mathrm{sh}^Y](x) \leq D_+ \mathbb{E}_F[\mathrm{sh}^Y](x)$ by the definition of $F(x)$ and $F(x-)$. This also shows that if F does not have a point mass at x, we have

$$\frac{d}{dx} \mathbb{E}_F[(Y-x)_+] = F(x) - 1, \tag{11}$$

coinciding with the "Leibniz rule" calculation

$$
\begin{aligned}
\frac{d}{dx} \mathbb{E}_F[(Y-x)_+] &= \frac{d}{dx} \int_x^\infty (y-x) f_Y(y) dy && (12)\\
&= \int_x^\infty \frac{d}{dx} (y-x) f_Y(y) dy - (x-x) f_Y(x) = -\int_x^\infty f_Y(y) dy = F(x) - 1. && (13)
\end{aligned}
$$

which assumes Y has an adequately well-behaved density f_Y.

3 Quantiles and Expected Shortage

We recall how quantiles arise as solutions to a probabilistic decision problem, drawing on perspectives developed in the fields of both forecasting (see e.g., Gneiting [2011a] and Jose and Winkler [2009]) and stochastic optimization (se e.g., Royset and Wets [2022], sections 1.C and 3.C).

Let Y be a random variable representing the future level of an undesirable outcome such as severe COVID incidence. Let $x \in \mathbb{R}_+$ be a decision variable representing levels of some costly counter-measure, such as procurement of monoclonal antibody treatments, that can be taken at a cost $C > 0$ per unit in preparation for Y.[2] A decision maker must decide on a level x of investment in the counter-measure, and wishes to avoid excesses in either the expediture Cx or the shortage $(y-x)_+$ when $Y = y$ is realized. To formalize the trade-off between these potential excesses we quantify the loss associated with a unit of shortage by a constant $L > C$ (which assumes that the counter-measure has some practical value) and combine the total shortage loss with expenditure into a *loss function*[3]

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

[2] Quantiles could also be derived for a problem in which x and Y take negative values, corresponding, for instance, to a decision maker that both buys and sells in a resource market and a Y that takes negative values when "recoveries" outnumber incidence. But we do not consider such scenarios in this work.

[3] This does involve a confusing use of the word *loss* to refer to two different quantities, but this seems to be an ingrained and unavoidable habit in the literature.

The decision problem is then to select a random future loss $l(x, Y)$ in a way that aligns with the preference that $l(x, y)$ be as low as possible given any realization $Y = y$.

To give the decision problem more structure we assume the decision maker either knows the distribution F of Y, or wishes to proceed as if a forecast F of Y were true. This gives us what is known in decision theory as a decision problem *under risk* (regarding the future value of Y) as opposed to one *under uncertainty* where both Y as well as F are unknown when the decision is to be made. A principle commonly invoked in this situation[4] is that the decision maker should or will seek to minimize the expected loss

$$\mathbb{E}_F[l(x, Y)] = Cx + L\mathbb{E}_F[(Y - x)_+]. \tag{14}$$

The expected loss is an affine transformation of the convex expected shortage (c.f. (1)). Therefore $\mathbb{E}_F[l(x, Y)]$ is also is convex and has right and left derivatives $D_\pm\mathbb{E}_F[l(x, Y)]$ at every x. Because these derivatives exist everywhere, a necessary condition for $x^\star$ to minimize $\mathbb{E}_F[l(x, Y)]$ is that $D_+\mathbb{E}_F[l(x^\star, Y)] \geq 0$ and $D_-\mathbb{E}_F[l(x^\star, Y)] \leq 0$, and because of convexity, this condition is also sufficient. From (6) and (10) this means that

$$D_+\mathbb{E}_F[l(x^\star, Y)] = C + L(F(x^\star) - 1) \geq 0 \geq D_-\mathbb{E}_F[l(x^\star, Y)] = C + L(F(x^\star-) - 1) \tag{15}$$

which rearranges with $\alpha = 1 - C/L$ to

$$F(x^\star) \geq \alpha \geq F(x^\star-). \tag{16}$$

Note that because $F(x)$ and $F(x-)$ are right and left continuous, repectively, and $0 < \alpha < 1$, the set $\{x \mid F(x) \geq \alpha\}$ is closed on the left and the set $\{x \mid \alpha \geq F(x-)\}$ is closed on the right. Therefore, (16) implies that

$$\min\{x \mid F(x) \geq \alpha\} \leq x^\star \leq \max\{x \mid \alpha \geq F(x-)\}. \tag{17}$$

We call $q^-_{\alpha,F} := \min\{x \mid F(x) \geq \alpha\}$ and $q^+_{\alpha,F} := \max\{x \mid F(x-) \leq \alpha\}$ the left and right quantiles of F (for probability level α) and any element $q_{\alpha,F} \in [q^-_{\alpha,F}, q^+_{\alpha,F}]$ a quantile of F. The *quantile function* for F, which we write as either $Q_F(\alpha)$ or $F^{-1}(\alpha)$, is the *set-valued* function that maps $\alpha \in (0, 1)$ to to the set $[q^-_{\alpha,F}, q^+_{\alpha,F}]$. Thus $x^\star$ minimizes the expected loss (14) and gives an optimal solution to the decision problem if and only if $x^\star \in Q_F(\alpha)$.

3.1 Quantile functions

For future reference, we record several key properties of quantile functions. An illuminating source for these facts is Rockafellar and Royset [2014], from which figure 2 is adapted.

The probability levels $\{\alpha_i\}$ for which $\#Q_F(\alpha_i) > 1$ ($\{\alpha_1, \alpha_5\}$ in the example of figure 2) form a discrete subset of $(0, 1)$ and correspond to the non-zero width intervals $[q^-_{\alpha_i,F}, q^+_{\alpha_i,F}]$ where F is constant with values $\{\alpha_i\}$. Conversely, if F is strictly increasing on a (Borel) set $A \subset \mathbb{R}$, then the restriction $Q_F\big|_{F(A)} = F^{-1}\big|_{F(A)}$ of Q_F to $F(A)$ is in fact a real-valued function which is left-continuous and provides an inverse to F on $F(A)$. If the the support $\text{supp}(F) = \{x \mid 0 < F(x) < 1\}$ of F is such an A, then extending F^{-1} by $F^{-1}(0) = \inf(\text{supp}(F))$ and $F^{-1}(1) = \sup(\text{supp}(F))$ provides an inverse to F on $F(\mathbb{R}) \subset [0, 1]$. If F has a point mass at $x \in A$ (e.g., $x_3 \in [0, \infty)$ in figure 2), so that $\{\alpha \mid F(x-) < \alpha < F(x)\}$ (e.g., (α_3, α_5) in figure 2) is a non-empty set disjoint from $F(A)$, then Q_F takes the constant value x on the closure $\{\alpha \mid F(x-) \leq \alpha \leq F(x)\}$. Conversely, if F has no discrete component on A, then $Q_F\big|_{F(A)}$ is strictly increasing (e.g., on $[\alpha_4, \infty)$ in figure 2).

Moreover, it can be said that Q_F is increasing as a set-valued function on $\mathbb{R}$ in the generalized sense that $(q_{\alpha,F} - q_{\beta,F})(\alpha - \beta) \geq 0$ whenever $q_{\alpha,F} \in Q_F(\alpha)$ and $q_{\beta,F} \in Q_F(\beta)$, that is, the set $\text{graph}(Q_F) = \{(\alpha, q) \mid q \in Q_F(\alpha)\}$ has no downward sloping secants. Conversely, given such an increasing set-valued function Q on $(0, 1)$, we can construct a right-continuous function F_Q from $\mathbb{R}$ to $[0, 1]$ which will be the cdf of the random variable $Y_Q := \min(Q(U))$ where $U \sim \text{Unif}[0, 1]$.

[4]Note that this priciple might be inappropriate when the decision maker is *risk averse* in some way such as having a preference for random losses with lower variance.

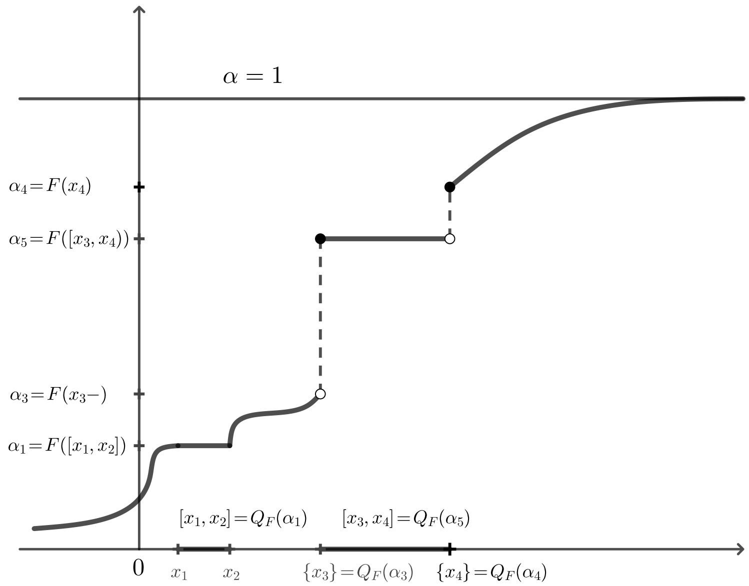

Figure 2: Example of a cdf F (with graph including solid line and filled points) superimposed on its quantile function Q_F (with graph including dashed and solid lines as well as both filled and empty points). The notation $F(A)$ should be read as $F(x)$ for any $x \in A$.

3.2 Opportunity relative to an oracle

Quantiles equivalently arise when the decision problem is defined in terms of the random *opportunity* loss

$$l_o(x, Y) := l(x, Y) - l(Y, Y) = Cx + L(Y - x)_+ - CY \tag{18}$$

which expresses how much more loss is realized by the decision x than an oracle would have incurred, knowing to invest exactly the future value of Y. The optimal decision for $\mathbb{E}_F[l_o(x, Y)]$ is the same as for $\mathbb{E}_F[l(x, Y)]$ since the term $-C\,\mathbb{E}_F[Y]$ is constant in x, leading again to the inequalities (15).

Opportunity loss (18) rearranges to

$$\begin{aligned} l_o(x, Y) &= C(x - Y)_+ + (L - C)(Y - x)_+ & (19)\\ &= L(1 - \alpha)(x - Y)_+ + L\alpha(Y - x)_+ & (20)\\ &= L(\alpha - \mathbf{1}\{Y < x\})(Y - x), & (21) \end{aligned}$$

a form in which it is often called *pinball* loss, despite its graph being an unlikely pinball trajectory for $\alpha \neq 1/2$.

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

For an arbitrary allocation vector $x \in \mathbb{R}_+^N$ the expected loss

$$\mathbb{E}_F[s_A(x,Y)] = \sum_{i=1}^{N} L \cdot \mathbb{E}_{F_i}[(Y_i - x_i)_+] \tag{22}$$

is the sum of expected shortages (scaled by L) under the allocations x_i in each location. We therefore have the following necessary condition for $x^\star \in \mathbb{R}_+^N$ to be an optimal allocation for $\mathbb{E}_F[s_A(x,Y)]$ under the constraint $\sum_{i=1}^N x_i = K$: if $\delta > 0$ of the $x_i^\star$ units of resource allocated to location i are reallocated to location j, expected shortage will increase in location i by at least as much as it decreases in location j. That is,

$$\mathbb{E}_{F_i}[(Y_i - (x_i^\star - \delta))_+] - \mathbb{E}_{F_i}[(Y_i - x_i^\star)_+] \geq \mathbb{E}_{F_j}[(Y_j - x_j^\star)_+] - \mathbb{E}_{F_j}[(Y_j - (x_j^\star + \delta))_+]. \tag{23}$$

Since the expected shortages in i and j have right and left derivatives at any x_i and x_j (see Section 2), we can divide (23) by δ and take limits for $\delta \searrow 0$ to get

$$-D_-\mathbb{E}_F[(Y_i - x_i^\star)_+] \geq -D_+\mathbb{E}_F[(Y_j - x_j^\star)_+]. \tag{24}$$

Note that the minus signs appear because our optimality condition addresses how a *decrease* in resources will *increase* the expected shortage in i and vice versa in j. Scaling by L to match the right and left partial derivatives of $\mathbb{E}_F[s_A(x,Y)]$ and using formulae (6) and (10), (24) becomes

$$L(1 - F_i(x_i^\star -)) \geq L(1 - F_j(x_j^\star)). \tag{25}$$

Inequalities (24) and (25) remain true with i and j reversed. They hold with $i = j$ as well by the definition of $F_i(x_i^\star -)$. Therefore, a number λ (a *Lagrange multiplier*) exists, which is independent of i, such that

$$L(1 - F_i(x_i^\star -)) \geq \lambda \geq L(1 - F_i(x_i^\star)), \quad \text{for all } i \in 1, \ldots, N. \tag{26}$$

That is,

$$F_i(x_i^\star) \geq 1 - \lambda/L \geq F_i(x_i^\star -), \tag{27}$$

which says (c.f. discussion after (16) and (17)) that $x_i^\star$ is a quantile q_{τ,F_i} for $\tau = 1 - \lambda/L$.

The constraint now determines τ (and hence the Bayes act) through

$$\sum_{i=1}^{N} q_{\tau,F_i} = K. \tag{28}$$

This equation implies that

$$K \in TQ_F(\tau) \tag{29}$$

where the set-valued function

$$TQ_F(\tau) := \sum_{i=1}^{N} Q_{F_i}(\tau) = \left[\sum_{i=1}^{N} q^-_{\tau,F_i}, \sum_{i=1}^{N} q^+_{\tau,F_i}\right] \tag{30}$$

is defined using interval addition $[a,b] + [c,d] = [a+c, b+d]$. (Note that the letter T is being used to connote a “totalling” operation, as it often is in survey sampling literature.) Conversely, if $\tau^\star$ satisfies (29), then we can find a solution $\boldsymbol{q}_{\tau^\star,F} := (q_{\tau^\star,F_1}, \ldots, q_{\tau^\star,F_N})$ to (28) within the set $\underset{\sim}{\boldsymbol{F}}^{-1}(\tau^\star) \subset \mathbb{R}_+^N$ where $\underset{\sim}{\boldsymbol{F}}^{-1}(\tau)$ is the Cartesian product $(F_1^{-1}(\tau), \ldots, F_N^{-1}(\tau))$.

TQ_F satisfies the conditions mentioned in section 3.1 for being a quantile function, and so there is a random variable TY_F with cdf $F_T := F_{TQ_F}$. From this perspective, the problem of finding τ becomes the calculation of $F_T(K) = \mathbb{P}(TY_F \leq K)$, making clear the existence of a solution $\tau^\star$ to (29). This also yields the interesting formal equation for the Bayes act

$$\underset{\sim}{\boldsymbol{F}}(x^{F,K}) = \mathbf{1}_N F_T(K), \tag{31}$$

where $\mathbf{1}_N : \mathbb{R} \to \mathbb{R}^N$ is the linear map that takes a to the N-vector $(a, \ldots, a)^T$ and the vector of marginal cdfs $\underset{\sim}{\boldsymbol{F}} := (F_1, \ldots, F_N)$ is a map from $\mathbb{R}^N$ to $\mathbb{R}^N$. With this notation we can write (29) as

$$K \in TQ_F\left(\frac{1}{N}\mathbf{1}^T \underset{\sim}{\boldsymbol{F}}(x^{F,K})\right), \tag{32}$$

which leads conceptually to the iterative numerical method of solving for τ and $x^{F,K}$ discussed next in section 5.

Two awkward features of the quantile representation of the Bayes act can arise. First, point masses in the F_i create point masses for F_T which may cause $\tau^\star$ to not be the unique solution to (29). Secondly, if more than one $Q_{F_i}(\tau^\star)$ is a positive-width interval, then the Bayes act will not be unique in these coordinates, and generically not all points in the $Q_{F_i}(\tau^\star)$ will be the coordinate of a Bayes act.

It is important to note that λ depends on the forecast F and the constraint level K. Thus while $\lambda = L(1-\tau)$ can be interpreted as a kind of "cost" imposed by the constraint in the allocation problem which is analogous to $C = L(1-\alpha)$ in the the cost-lost problem of section 3, it does not serve to define the allocation loss function in the way that C defines (14). λ is rather a parameter that must be found, given the pair F and K.

5 Numerical computation of allocation scores

Exact computation of the allocation score $S_A(F, y; K)$ of a forecast F requires an explicit choice of and expression for a solution $x^{F,K} = \boldsymbol{q}_{\tau^\star,F} \in \underset{\sim}{\boldsymbol{F}}^{-1}(\tau^\star)$ to (28) where $\tau^\star$ solves (29). For practical scenarios in which the forecast F has a form preventing analytical solutions to (28), we seek a method of generating an approximation $\tilde{\tau}$ to a solution $\tau^\star$ which offers control (at least in principle) over a robust measure of scoring rule error such as the sum of absolute coordinate errors

$$\mathrm{SAE}(Y, \tilde{\tau}) = \sum_{i=1}^{N} \left|(Y_i - q_{\tilde{\tau},F_i})_+ - (Y_i - x_i^{F,K})_+\right| \tag{33}$$

where $\boldsymbol{q}_{\tilde{\tau},F}$ solves the linear program $\sum_{x \in \underset{\sim}{\boldsymbol{F}}^{-1}(\tilde{\tau})} x_i = K$.

Suppose we have established that $\tau^\star = F_T(K)$ 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 3.1, 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 2):

(●●) $TQ_F(\tau_L) \cap TQ_F(\tau_U) = \varnothing$ (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., $\alpha_3 \le \tau_L \le \tau_U \le \alpha_5$)

(●–) $TQ_F(\tau_L) = \{q^-_{F_T,\tau_U}\} \subsetneq TQ_F(\tau_U)$ (a point mass at $q^-_{F_T,\tau_U}$, 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 $q^+_{F_T,\tau_U}$, e.g., e.g., $\tau_L = \alpha_5, \tau_U = \alpha_4$).

In the case (●), we can immediately take $\tau^\star = \tau_U$ 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 $\tau^\star = \tau_U$ or τ_L, respectively, as the representing probability level. Having found our $\tau^\star$ we then arrive at a Bayes act $x^{F,K}$ by by solving the linear program $\sum_{x \in \underset{\sim}{\boldsymbol{F}}^{-1}(\tau^\star)} x_i = K$.

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

$$I_2 = \begin{cases} [\tau_L, \tau_M] & \text{if } K < q^-_{F_T,\tau_M} \\ [\tau_M, \tau_U] & \text{if } K \ge q^-_{F_T,\tau_M} \end{cases} \tag{34}$$

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

$$\begin{cases} K < q^-_{F_T,\tau_M} \text{ implies } \tau^\star = F_T(K) < \tau_M \\ K \ge q^-_{F_T,\tau_M} \text{ implies } \tau^\star = F_T(K) \ge \tau_M. \end{cases} \tag{35}$$

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 $\{I_k\}, k = 1, 2, \ldots$ 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 $\tau^\star$ from below and above. Such a sequence provides the basis of a "bisection" algorithm for finding the "root" $\tau^\star$ 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 2 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 $\sum_{x \in \underset{\sim}{F}^{-1}(I_K)} x_i = K$ 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 $\mathcal{T} = \{0.01, 0.025, 0.05, 0.1, 0.15, \ldots, 0.9, 0.95, 0.975, 0.99\}$, 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 the allocation score

In order for a scoring rule to be guaranteed to yield logically consistent forecast evaluations, it must be *proper*, which means, in essence, that a forecaster cannot perceive any forecast that is at odds with their beliefs about the events being forecasted as having an optimal expected score. We formalize this in terms of forecaster "self-assessment" in section 7.1 and then address the propriety of the allocation score in two stages:

- Section 7.2 shows that the allocation score is proper when the forecaster's actual forecast distribution F is available (as is any score derived using the decision theoretic procedures outlined in section 2 of the manuscript)
- Section 7.3 addresses the propriety of an analysis that combines an approximate parametric representation of the forecast distribution and evaluation with the allocation score. It shows that this procedure yields a proper score so long as this representation and scoring procedure was specified prospectively by the hub. This includes the use of methods such as `distfromq` for constructing forecast distributions as a special case. However, *post hoc* application of this scoring procedure to quantile forecasts is improper.

7.1 Proper scores

Expanding on the decision theoretic perspective sketched in Section 3, we can view a loss function as a general tool for formalizing a decision problem that assigns numerical value to the *result* of taking an *action* x in preparation for an *outcome* y. Here, we consider a setting where the individual making the decision is the forecaster, and the action is to report a particular forecast.

A *scoring rule* S is a loss function where the action is a probabilistic forecast F of the outcome y (or the statement of F by a forecaster). Just as in Section 3, given an action F, S transforms a random outcome variable Y into a random loss $S(F, Y)$. 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. S therefore commits a probabilistic forecast F to the "self-assessment" $\mathbb{E}[S(F, Y^F)]$, where $Y^F \sim F$ is the random variable defined by sampling from F, as well to an assessment $\mathbb{E}[S(G, Y^F)]$ 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

$$\mathbb{E}[S(F, Y^F)] \leq \mathbb{E}[S(G, Y^F)] \tag{36}$$

for any F, G. Otherwise, the optimal decision for some forecaster would be to state a forecast G other than the forecast F which they believe describes the outcome Y. A scoring rule meeting this criterion is called *proper*. The inequality can also be written as $\mathbb{E}_F[S(F, Y)] \leq \mathbb{E}_F[S(G, Y)]$ where the subscript specifies the distribution of Y. S is *strictly proper* when this inequality is sharp, in which case the *only* optimal decision for a forecaster is to state the forecast they believe to be true.

7.2 The allocation score is proper

We recall the three-step decision theoretic procedure for deriving proper scoring rules that we outlined in section **??** of the main text:

1. Specify a *loss function* $s(x, y)$ that measures the loss associated with taking action x when outcome y eventually occurs.
2. Given a probabilistic forecast F, determine the *Bayes act* x^F that minimizes the expected loss under the distribution F.
3. Define a *scoring rule* which assigns the loss realized by the Bayes act x^F against outcome y as the score of F at y:

$$S(F, y) = s(x^F, y). \tag{37}$$

Such scoring rules, which we call *Bayes scoring rules*, are proper by construction since

$$\begin{aligned}
\mathbb{E}_F[S(F,Y)] &= \mathbb{E}_F[s(x^F,Y)] \\
&= \min_x \mathbb{E}_F[s(x,Y)] \quad \text{(by definition of } x^F\text{)} \qquad (38) \\
&\leq \mathbb{E}_F[s(x^G,Y)] \qquad (39) \\
&= \mathbb{E}_F[S(G,Y)].
\end{aligned}$$

The allocation scoring rule is Bayes and therefore proper.

We note that in the probabilistic forecasting literature (see e.g., Gneiting [2011b], Theorem 3) what we have termed Bayes scoring rules typically appear via (37) 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.

7.3 Prospectively-specified use of the allocation score to evaluate forecasts approximately represented as members of a parametric family is proper

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 consitutes 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 Wadsworth et al. [2023]. Additionally, we note that the functionality in `distfromq` can be viewed as specifying a parametric family $\mathcal{P}_{\text{dfq}}$ 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.

We find it helpful now to formally distinguish between two decision making problems. The first is the public health decision maker's allocation problem where the task is to select an allocation x, with the allocation loss $s_A(x,y) = \sum_{i=1}^N L \cdot \max(0, y_i - x_i)$ as described in section **??** of the main text. The second is the forecaster's reporting problem where the task is to select parameter values θ to report. The forecaster's loss is given by $s_R(\theta, y) = s_A(x^{G_\theta}, y)$, where x^{G_θ} is the Bayes act for the allocation problem under the distribution G_θ. In words, the loss associated with reporting θ is equal to the loss associated with taking the Bayes allocation corresponding to the distribution G_θ.

Following our usual construction, the Bayes act for the forecast reporting problem is the parameter set that minimizes the forecaster's expected loss. Breaking with our earlier notation for improved legibility, we use $\theta^\star(F)$ to denote this Bayes act:

$$\begin{aligned}
\theta^\star(F) &= \text{argmin}_\theta \, \mathbb{E}_F[s_R(\theta, Y)] \\
&= \text{argmin}_\theta \, \mathbb{E}_F[s_A(x^{G_\theta}, Y)]
\end{aligned}$$

We then arrive at the scoring rule

$$S_R(F, y) = s_R(\theta^\star(F), y) = s_A(x^{G_{\theta^\star(F)}}, y).$$

It follows from the discussion in section 7.2 that this is a proper scoring rule for F. Additionally, we note that the score can be calculated from the reported parameter values $\theta^\star(F)$.

We emphasize that the forecaster's true predictive distribution F does not need to be a member of the specified parametric family $\mathcal{P}$ for this construction to yield a proper score. It is, however, necessary to specify the parametric family to use and the foundational scoring rule s_A (including any relevant problem parameters such as the resource constraint K) in advance, so that forecasters can identify the Bayes act parameter set $\theta^\star(F)$ to report.

If the parametric family used to represent forecast distributions is flexible enough, the reporting scoring rule S_R and the allocation score are equivalent in the sense that they will yield the same score for any distribution F. Suppose that for a given resource constraint K, for any forecast distribution F it is possible to find a member $G_{\theta^\star}$ of the specified parametric family $\mathcal{P}$ with the same allocation as F (i.e., $x^F = x^{G_{\theta^\star}}$). Then $\theta^\star$ is a Bayes act for the reporting problem since for any other parameter value θ,

$$\begin{aligned}
\mathbb{E}_F[s_R(\theta^\star, Y)] &= \mathbb{E}_F[s_A(x^{G_{\theta^\star}}, Y)] \\
&= \mathbb{E}_F[s_A(x^F, Y)] \quad \text{(since } x^F = x^{G_{\theta^\star}}\text{)} \\
&\leq \mathbb{E}_F[s_A(x^{G_\theta}, Y)] \quad \text{(by definition of } x^F\text{)} \\
&= \mathbb{E}_F[s_R(\theta, Y)].
\end{aligned}$$

It therefore follows that

$$\begin{aligned}
S_R(F, y) &= s_R(\theta^\star, y) \\
&= s_A(x^{G_{\theta^\star}}, y) \\
&= s_A(x^F, y) \\
&= S_A(F, y).
\end{aligned}$$

For the particular choice of the parametric family $\mathcal{P}_{\text{dfq}}$ (i.e., using the `distfromq` package), this flexibility requirement is satisfied. For instance, the forecaster could pick one required quantile level (such as 0.5, for which the corresponding predictions are predictive medians), and set the submitted quantiles of their forecast distribution at that level to be the desired allocations. However, this representation of the forecast may be quite different from the actual forecast distribution F.

A key observation is that a *post hoc* analysis combining `distfromq` with the allocation score does not in general yield a proper evaluation method. This is because the forecast distribution F and the distribution $G_\theta \in \mathcal{P}_{\text{dfq}}$ with the same quantiles as F may determine quite different resource allocations, particularly if the tail extrapolations performed by `distfromq` do not match the tail behavior of F.

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.

References

Tilmann Gneiting. Quantiles as optimal point forecasts. *International Journal of forecasting*, 27(2): 197–207, 2011a.

Victor Richmond R Jose and Robert L Winkler. Evaluating quantile assessments. *Operations research*, 57(5):1287–1297, 2009.

Johannes O Royset and Roger J-B Wets. *An optimization primer*. Springer, 2022.

R Tyrrell Rockafellar and Johannes O Royset. Random variables, monotone relations, and convex analysis. *Mathematical Programming*, 148:297–331, 2014.

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

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

Spencer Wadsworth, Jarad Niemi, and Nick Reich. Mixture distributions for probabilistic forecasts of disease outbreaks. *arXiv preprint arXiv:2310.11939*, 2023.