Allocation scores, WIS, and CRPS via decision theory

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

December 9, 2023

Abstract

We develop a unified decision theoretic framwork for quantile scores, WIS, CRPS, and allocation scores.

1 Introduction

In this report, we present a methodological perspective on the use scoring rules for forecast evaluation that we see as well-adapted to our motivating research goal to understand and describe the social utility of infectious disease forecasts. Our approach follows the common practice in decision theory of giving central roles to the foundational concepts of state, action, and result as the basic components of a decision problem. From these primitives, we first derive the familiar measures of quantile score (QS), continuous ranked probability score (CRPS) and their aggregates across forecast targets. We then introduce new scoring rules, the *specific* and *integrated allocation scores* designed to reflect aspects of the decision problem beyond those defining the QR and CRPS that emerge when multiple instances of a decision problem are confronted simultaneously. Specifically, we will consider not only whether using more or less of a costly resource leads to a better result in a specific problem, but also whether a limited resource is better allocated to one problem rather than another. This added concern brings into play the duality in mathematical optimization theory between constraints on decision variables on the one hand and dual or "Lagrange multiplier" variables on the other. Dual variables quantify the sensitivity of the optimal value of an objective function to constraints, that is, the degree to which loosening constraints improves results under the optimal action. Our computational methods (see ..) in fact make use of this connection. Something about Bayes risks...

2 Scoring forecasts via decision problems

Broadly speaking, we view scoring a probabilistic forecast F with a scoring rule as the process of identifying F with a solution to a *decision problem* and then rating the utility of that solution. The term "decision problem" refers to a pre-specified set of allowed actions, possible future states, and the possible results of allowed actions under those states. Accordingly, F will receive a "better" or "worse" score in a particular instance when the result is better or worse of a rational decision maker using F (and essentially only F) to select an action in the face of uncertainty about a future state. As such scores accumulate, we can use their empirical distributions across forecasting instances as inputs to higher level decision problems: whether a status quo policy should be adjusted according to the forecasts produced by a particular model or method; whether one forecasting method is a better choice to guide policy rather than another; how different forecasting methods can best be coordinated or combined to guide policy. By connecting such policies with the specific decision problems we use to define scoring rules, we may be able to improve these higher level decisions we make about how forecasts are used.

2.1 The ingredients of a decision problem with examples

The basic example of a future state we are considering is the number of individuals that become sick enough from an infectious disease to be in need of some potentially limited resource such as hospital beds, ventilators, medication, or medical staff. We write y to refer to a particular state such as the quantity of need in a single location or the quantities across multiple locations or times and Y for a

future uncertain state which is a random variable with a set of realizable values $\mathcal{Y}$. The forecast F to be scored is a claim of how Y is (or will be) distributed. We also will refer to states as outcomes, which helps to emphasize their future realization. But we want to clearly distinguish an outcome from a result, which as we discuss below, is determined by an action being taken under a certain outcome.

The basic example of an action we consider is the specific amount or set of amounts of such a resource that a public health worker decides to make available in either a single location or across multiple locations or times. We write an action as x and assume that it takes values in a set of possible actions $\mathcal{X}$ that is possibly defined by some constraint such as the total amount of resource available to the public health worker for allocation. Note that x must be selected prior to the realization of y. A decision maker might use a set rule to select x based on inputs, such as F, but the observed value of y cannot be such an input.[1] We will also require that y be observed before x can effect y. Thus, if we were considering actions of providing vaccines or masks to protect a (likely estimated) number y_e of people exposed to an observed number y of infectious individuals, we would require y be observed before the vaccines or masks had any chance to reduce disease incidence.

And when actions are assigned to states, there are results, which we abstractly identify with an element r of a set $\mathcal{R}$. We have two basic examples in mind of what such a r might be. First, we consider some amount of the regret mentioned above that resources made available in a place or time are not actually needed at that place or time. We emphasize that the nature of this regret is potentially quite amorphous. In business contexts (e.g., the "newsvendor problem" from inventory management) this regret is often easily quantifiable as monetary waste, but in a public health context, it could... (A desire to better understand over-allocation regret and its relevance to infectious disease forecast evaluation was actually a primary impetus for the present work.) Our second and more concrete example of a result is the loss incurred when a sick individual is not able to access a resource because it was not made available to them where and when they needed it. We emphasize here that while we formally identify such loss as a result of a particular action being taken under a particular state, it may or may not be an avoidable loss. That is, it could just as well result from a factor outside a decision maker's control such as a severe resource constraint as from a choice to withhold a resource, motivated, for example, by a desire to reduce waste.

2.2 The formal procedure

With a view towards a decision problem defined by these examples of sets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{R}$ of actions, outcomes, and results, we now give a formal three-step procedure for defining and evaluating a *scoring rule* for a forecast F for a future value $y \in \mathcal{Y}$ of Y.

1. Specify a *loss function* $s(x, y)$ that quantifies the utility or disutility of the result r of taking action x under outcome y.

2. Map the given probabilistic forecast F for Y to an action x^F which "solves" the decision problem by minimizing the expected loss $E_F[s(x, Y)]$ over all $x \in \mathcal{X}$ under the distribution F. Following Dawid [2007], we call x^F the *Bayes act* for F. Note that x^F depends not only on F and s, but also on the definition of $\mathcal{X}$. For example, having more available resources usually allows for actions incurring less loss.

3. Assign F the score $s(x^F, y)$, that is, the numerical quantification via s of the result under outcome y of accepting x^F as a solution to the decision problem.

Thus, given definitions of $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{C}$, and a loss function s, we produce a scoring rule $S(F, y) := s(x^F, y)$ that can be applied to any F for any outcome $y \in \mathcal{Y}$. We emphasize again a key feature of this formulation: the scoring rule S is defined entirely in terms of what we take to be possible actions, states, and results along with a function s that assigns values to the possible results.

Another key feature of this procedure is that the scoring rule S it yields is proper by definition. This is because a forecaster that believes that Y has distribution F and understands the decision problem and loss function s must also believe that x^F is the optimal action to take. This, in turn, implies the belief that there is no forecast that will give a better expected score than F. Therefore, there is no incentive to reveal a forecast at odds with a forecaster's beliefs about Y.

---

[1] In a dynamical decision theoretic approach, the action x could be a time series depending at each step on previous values of y. This could be a very fruitful framework for infectious disease forecast evaluation but we do not pursue it is this work.

Remark. For there to be an incentive *not* to reveal a forecast other than F we would need a *strictly* proper scoring rule, which would require more conditions on the decision problem and/or s. In the literature, a functional $M(F)$ (such as the mean or median of F) is called *elicitable* if there is a strictly proper scoring rule $S(F, y) = s(x^F, y)$ for which $M(F) = x^F$. But the focus in work on elicitability often seems to be the functional T as a starting point rather than the decision problem, which is our primary concern here.

The following sections explain in detail how the QS, WIS, CRPS (for various parameters and weightings) as well as extensions to higher dimnsional setting can be seen as outputs of our procedure.

3   The quantile score

We begin by imagining a public health worker trying to decide at time t_0 what quantity x to procure of a protective resource such as beds, staff, or medicine in anticipation of y new infections that will be observed (and create resource demand) at time $t_1 > t_0$. The resource has a unit-cost of C, and we assign a loss of L to the result that a sick individual cannot access the resource. We assume that $C < L$ since otherwise – at least according to the quantification L – there is no "cost-effectiveness" in trying to prevent even a single case of unmet need. With C and L, there are two natural loss function we can define. First, we have the *absolute* loss

$$s_{Q,C,a}(x, y) = Cx + L(y - x)_+ \tag{1}$$

where $u_+ = \max(u, 0)$. Alternatively, we have the *opportunity* loss

$$s_{Q,C,o}(x, y) = Cx + L(y - x)_+ - Cy \tag{2}$$
$$= C(x - y)_+ + (L - C)(y - x)_+ \tag{3}$$
$$= (C - L\mathbf{1}\{y > x\})(x - y) \tag{4}$$
$$= L(\mathbf{1}\{y \le x\} - 1 + C/L)(x - y) \tag{5}$$

which considers only the loss relative to that of an oracle that is able to minimize loss *ex-post* by procuring exactly y units of the resource.

Loss functions of these general forms are referred to as *piece-wise linear*, distinguishing them from smooth non-linear loss functions such as squared error. The opportunity loss $s_{Q,C,o}$ is sometimes called "pinball" loss. It has the convenient property that $s_{Q,C,o}(z, z) = 0$, but loses some of the interpretability of $s_{Q,C}$ (see **??** below). The underlying decision problem here consists of action and outcome sets $\mathcal{X}, \mathcal{Y}$ both equal to $\mathbb{R}_+$, the non-negative reals. (We ignore for now any discreteness of the resource.) The results set $\mathcal{R}$ is abstractly a set of transfers of a monetary or similar nature and sick individuals with unmet resource need, but the loss function s maps $\mathcal{R}$ into $\mathbb{R}_+$.

Note that the absolute and opportunity losses have the same dependence on action variable x, and therefore the same Bayes act. For $C \in (0, L)$ this is the solution of the first order equation

$$0 = \frac{d}{dx} E_F[s_{Q,C}(x, Y)] = E_F\left[\frac{d}{dx} s_{Q,C}(x, Y)\right] \tag{6}$$
$$= C + LE_F\left[\frac{d}{dx}(Y - x)_+\right] \tag{7}$$
$$= C - LE_F[\mathbf{1}\{Y > x\}] \tag{8}$$
$$= C + L(F(x) - 1), \tag{9}$$

that is, the quantile $x^F = F_i^{-1}(1 - C/L) = q_{F,\tau}$ for the probability level $\tau = 1 - C/L$. (That the critical point $q_{F,\tau}$ is actually a minimum follows from (9) being non-positive to its left and non-negtive to its right, and assuming as we do for now that F is smooth and strictly increasing at $q_{F,\tau}$, it is a unique minimum.) The quantile solves the decision problem with respect to the forecast F by being the action which best balances expected cost against expected loss among all actions available to the decision maker.

Finally, we evaluate the scoring rules $S_{Q,C,a}$ and $S_{Q,C,o}$ on F when outcome y realizes by measuring the loss that occurs when the action $q_{F,\tau}$ is taken:

$$S_{Q,C,a}(F, y) = s_{Q,C,a}(q_{F,\tau}, y) = Cq_{F,\tau} + L(y - q_{F,\tau})_+ \tag{10}$$
$$S_{Q,C,o}(F, y) = s_{Q,C,o}(q_{F,\tau}, y) = (C - L\mathbf{1}\{y > x\})(q_{F,\tau} - y). \tag{11}$$

4 WIS and CRPS as quantile scores under cost uncertainty

The decision problem leading to the QS may however be complicated by uncertainty at time t_0 about the cost C of the resource. We take such uncertainty as requiring the worker to commit in advance to procurement levels x_j for a range of potential costs $C_j < L$ where $j \in 1, \dots, m$. Even though only one of the x_i will end up needing to be procured, there does not seem to be any intrinsically decision theoretic principle by which our loss function should ignore counter-factual but exactly specified results of the decision under unrealized costs. If we wanted to commit to a *locality* principle for prediction where losses and scores only depend on predictive densities in a neighborhood of the relevant observation, this attempt would falter, and we would need to frame our methods around the log score and it's higher order relatives. But we do not see that as an appropriate principle in public health resource allocation where probabilistic forecasts can be valuable by pushing decision makers in the "right direction" even when they assign low probability to observed outcomes. We instead adopt a "sensitivity-to-distance" principle from which piece-wise linear loss functions and the QS can be shown to naturally emerge. It is this principle that makes the loss under alternative costs seem relevant. **APG:[These are obviously incomplete thoughts but I feel like there is something here that I've been grasping for for a lng time.]**

We take into account the results of actions under a set $\mathcal{C} = \{C_j\}$ of possible costs by expanding (either the absolute or opportunity version of) the loss function to be a weighted mean of quantile losses under certainty about the cost:

$$s_{Q,\mathcal{C},p_C}(x, y) = \sum_{j=1}^{m} p_j s_{Q,C_j}(x_j, y). \tag{12}$$

Here, $p_C = \{p_j\}, j = 1, \dots, m$ is a prior distribution on $\mathcal{C}$ and $x = (x_1, \dots, x_m)$ is the decision vector corresponding to all possible costs. p_C would be provided by the decision maker (or perhaps a separate expert), and we emphasize that we are considering it here as part of the loss function $s_{Q,\mathcal{C},p_C}$ that would be formulated in an effort to adapt forecast evaluation to the specifics of a public health decision problem. In practice, p_C has usually been taken to be uniform, which we will show for a particular $\mathcal{C}$ (given L) leads to the WIS familiar in ID forecasting hubs. This might be interpreted as deferring to an uninformative prior.

We can also consider C a continuous random variable with a prior density f_C and accommodate this by letting $m \to \infty$ in (12) which (with some mathematical care) converts the weighted mean into an integral

$$s_{Q,f_C}(x, y) = \int_0^L s_{Q,c}(x(c), y) f_C(c) dc. \tag{13}$$

It is important to note that these generalizations change not only the loss function but also the structure of the decision problem by using higher dimensional action sets $\mathcal{X} = \mathbb{R}_+^m$ or the set $\mathcal{X} = \{x(c) : [0, \infty) \to \mathbb{R}_+\}$ of $\mathbb{R}_+$ valued functions on $[0, \infty)$.

The Bayes acts for losses $s_{Q,\mathcal{C},p_C}$ and s_{Q,f_C} therefore correspond not just to the vanishing of the derivative with respect to a single decision variable x but to the vanishing of the partial derivatives with respect to all m decision variables $\{x_i\}$ in the first case or the infinitely many variables $\{x(c)\}$ in the second case. But since the losses are sums of single variable losses, passing to solutions of these higher diminsional equations is a mere formality of the viewing the set of Bayes acts, i.e. quantiles, for costs $c \in \mathcal{C}$ or $(0, L)$ as a vector $(x^F(1), \dots, x^F(m)) = (q_{F,\tau_1}, \dots, q_{F,\tau_m})$ or a function $x(c) = q_{F,\tau(c)}$.

Accordingly, evaluation of the scoring rules $S_{Q,\mathcal{C},p_C}$ and S_{Q,f_C} on the forecast F given an outcome y amounts to summing the formula for $S_Q(F, y)$ across such a vector or function. That is,

$$S_{Q,\mathcal{C},p_C}(F, y) = \sum_{j=1}^{m} p_j s_{Q,C_j}(q_{F,\tau(C_j)}, y) \tag{14}$$

$$S_{Q,f_C}(F, y) = \int_0^L s_{Q,c}(q_{F,\tau(c)}, y) f_C(c) dc. \tag{15}$$

How do the WIS and CRPS arise from this construction? Defining $\tau = 1 - C/L$ as a random

variable gives it the density $f_\tau(\tau) = f_C(L(1-\tau)) \cdot L$, and (15) with the opportunity loss (5) becomes

$$S_{Q,f_C}(F,y) = \int_0^L L(\mathbf{1}\{y \le q_{F,\tau(c)}\} - \tau(c))(q_{F,\tau(c)} - y)f_C(c)dc \tag{16}$$

$$= \int_0^1 (\mathbf{1}\{y \le q_{F,\tau}\} - \tau)(q_{F,\tau} - y)f_\tau(\tau)d\tau \tag{17}$$

$$= w\text{CRPS}(F,y) \tag{18}$$

where $w\text{CRPS}$ is the weighted CRPS of Gneiting and Ranjan [2011] with quantile weight function $\nu(\tau) = \frac{1}{2}f_\tau(\tau)$. Using the absolute loss instead of opportunity loss would add a term $E[C]y$. This is the value of the absolute loss scoring rule at the oracle's point mass forecast: $S_{Q,f_C,a}(\delta_y, y)$.

To connect with the WIS, we recall it's definition by Bracher et al. [2021]:

$$\text{WIS}_{\alpha_{\{1:J\}}}(F,y) = \frac{1}{J+1/2}\left(w_0|y - q_{F,.5}| + \sum_{j=1}^J w_j\,\text{IS}_{\alpha_j}(F,y)\right). \tag{19}$$

The scoring rule $\text{IS}_\alpha(F,y)$ is the *interval score* suggested by expression (43) in Gneiting and Raftery [2007] and formalized by Bracher et al. [2021] as

$$\text{IS}_{\alpha_j}(F,y) = \frac{\alpha_j}{2}(q_{F,\alpha_j/2} - y)_+ + \frac{\alpha_j}{2}(y - q_{F,1-\alpha_j/2})_+ + (q_{F,1-\alpha_j/2} - q_{F,\alpha_j/2}) \tag{20}$$

Thus the α_j are twice the probability level of the lower endpoint of a central $1 - \alpha_j$ prediction interval. The COVID-19 Forecast Hub chooses $\alpha_j = 0.02, 0.05, 0.1, \ldots, 0.9, j = 1, \ldots, 11$.

Now write $s_{Q,\tau,o}(x,y) = s_{Q,C,o}(x,y)/L = (1-\tau)(x-y)_+ + \tau(y-x)_+$, and assume $\tau < 1/2$. We have

$$s_{Q,\tau,o}(q_{F,\tau},y) + s_{Q,1-\tau,o}(q_{F,1-\tau},y) \tag{21}$$

$$= (1-\tau)(q_{F,\tau} - y)_+ + \tau(y - q_{F,\tau})_+ + \tau(q_{F,1-\tau} - y)_+ + (1-\tau)(y - q_{F,1-\tau})_+ \tag{22}$$

$$= (q_{F,\tau} - y)_+ + (y - q_{F,1-\tau})_+ \tag{23}$$

$$\quad + \tau\left[(y - q_{F,\tau})_+ - (q_{F,\tau} - y)_+ + (q_{F,1-\tau} - y)_+ - (y - q_{F,1-\tau})_+\right] \tag{24}$$

$$= (q_{F,\tau} - y)_+ + (y - q_{F,1-\tau})_+ + \tau(q_{F,1-\tau} - q_{F,\tau}) \tag{25}$$

$$= \tau\text{IS}_{2\tau}(F,y). \tag{26}$$

To reproduce the WIS we set $\mathcal{C} = \{C_k\}, k = 0, \ldots 2J$ where

$$C_0 = L/2 \tag{27}$$

$$\alpha_k/2 = 1 - C_k/L, k = 1, ..J \tag{28}$$

$$1 - \alpha_{k-J}/2 = C_k/L, k = J+1, \ldots, 2J \tag{29}$$

(note that C_k decrease as k increases for $k > 0$) and let p_C be a probability distribution symmetric across $L/2$. This gives us

$$S_{Q,\mathcal{C},p_C,o}(F,y) = p_0 s_{Q,L/2,o}(q_{F,.5},y) + \sum_{k=1}^{2J} p_k s_{Q,C_k,o}(q_{F,1-C_k/L},y) \tag{30}$$

$$= p_0 s_{Q,L/2,o}(q_{F,.5},y) + \sum_{k=1}^J L(1 - C_k/L)p_k\text{IS}_{2(1-C_k/L)}(F,y) \tag{31}$$

$$= \frac{L}{2}\left(p_0|y - q_{F,.5}| + \sum_{j=1}^J \alpha_j p_j\text{IS}_{\alpha_j}(F,y)\right) \tag{32}$$

$$\propto \text{WIS}_{\alpha_{\{0:J\}}}(F,y) \tag{33}$$

if we set $p_j \propto w_j/\alpha_j$ and define $\alpha_0 = 1$. The COVID-19 Forecast Hub chooses $w_j = \alpha_j/2$ so that the WIS there corresponds to a uniform distribution over the C_k where each cost has probability

$$p = \frac{1/2}{\sum_{k=0}^{2J} 1/2} = \frac{1/2}{J+1/2}. \tag{34}$$

In this case the constant of proportionality in (33) is $L/2$.

Again, the hierarchy we have introduced here describing WIS and CRPS as parallels to the quantile score for higher dimensional action sets is in itself a mere formality employing the linearity of differentiation with respect to the expectation operator E_C for C considered as a random variable. But as we move now to decision problems with multiple outcomes, it will turn out to provide a key shift in perspective that allows the scoring rules we define in terms of allocation under scarcity to be seen as a natural and non-trivial extension of the hierarchy.

5   Aggregating QS, WIS, and CRPS across dimensions

But before introducing an allocation scoring rule, we need to expand our decision problem along one more dimension to include composite actions that address multivariate outcomes $y = (y_1, \ldots, y_N)$ being forecasted by a multivariate distribution F with marginal distributions $F_i, i = 1, \ldots, N$. This includes our motivating scenario of simultaneous procurement of medical countermeasures across several locations (U.S. states and territories) during a pandemic. And again, we simply aggregate the lower dimensional losses (1), (12), or (13) over the locations (or multiple forecast targets more generally) in order to formally obtain a single loss function that is reactive to all components of the overall procurement decision.

The action sets for the decision problem corresponding to a single certain cost C, a discrete cost distribution over $\{C_j\}, j = 1, \ldots, m$, and a continuous cost distribution f_C with support on $[0, L]$ are now

$$\mathcal{X}_{cert} = \mathbb{R}_+^N \tag{35}$$

$$\mathcal{X}_{disc} = (\mathbb{R}_+^m)^N = \{x(j) : \{1, \ldots, m\} \to \mathbb{R}_+^N\} \tag{36}$$

$$\mathcal{X}_{cont} = \{x(c) : [0, \infty) \to \mathbb{R}_+^N\}. \tag{37}$$

For these "vector-action" problems we define *total* versions of our loss functions

$$s_{TQ,C}(x, y) = \sum_{i=1}^N Cx_i + L(y_i - x_i)_+ \tag{38}$$

$$s_{TQ,\mathcal{C},p_C}(x, y) = \frac{1}{m} \sum_{i=1}^N \sum_{j=1}^m p_j(C_i x_{i,j} + L(x_{i,j} - y_i)_+) \tag{39}$$

$$s_{TQ,f_C}(x, y) = \sum_{i=1}^N \int_0^L (cx_i(c) + L(x_i(c) - y_i)_+) f_C(c) dc. \tag{40}$$

Similarly as for the expansion to multiple cost contingencies—though now with respect to forecasts F_i depending on the index i of the relevant outcome component y_i—first order conditions for optimality decouple under the additive separability of the loss function so that the Bayes act x^F for F is just a vector of quantiles $q_{F,\tau} := (q_{F_1,\tau}, \ldots, q_{F_N,\tau})$ or quantile functions when $\tau = 1 - c/L$, $c \in \mathcal{C}$ or $[0, L]$. Note that while $x^F = q_{F,\tau}$ only depends on the marginals F_i, it is indeed the Bayes act for the joint F since $E_F[s_{TQ,c}(x, y)] = \sum_{i=1}^N E_{F_i}[s_{Q,c}(x_i, y_i)]$.

Combining the total loss functions with these composite Bayes acts gives us total versions of scoring rules for the composite decisiion problem

$$S_{TQ,C}(F, y) = \sum_{i=1}^N Cq_{F_i,\tau} + L(y_i - x_i)_+ \tag{41}$$

$$S_{TQ,\mathcal{C},p_C}(F, y) = \frac{1}{m} \sum_{i=1}^N \sum_{j=1}^m p_j(C_j q_{F_i,\tau_j} + L(q_{F_i,\tau_j} - y_i)_+) \tag{42}$$

$$S_{TQ,f_C}(F, y) = \frac{1}{L} \sum_{i=1}^N \int_0^L (cq_{F_i,\tau(c)} + L(q_{F_i,\tau(c)} - y_i)_+) f_C(c) dc. \tag{43}$$

A key point for our presentation will be that the same C, set $\{C_j\}$, probabilities $\{p_j\}$, and density f_C are used for all locations simultaneously. We believe that our methods could be adapted to variations across locations (which could be important in practice), but we leave this generalization for later work.

6   The allocation score

The loss functions (and corresponding decision problems) defined so far have all been thouroughly covered, from one perspective or another, in various strands of forecast evaluation literature. We now introduce one that, while certainly familiar in operations research and related fields, has not, to our knowledge, been explicitly discussed in a forecast evaluation context.

Consider again the resource procurement decision $x(c) \in \mathcal{X}_{cont}$ (or $\mathcal{X}_{disc}$ if $c = C_j, j \in 1, \ldots, m$) a public health worker must make in attempting to mitigate a pandemic across multiple locations $i \in 1, \ldots, N$. But now we ask the worker to extend their decision to cover the contingency that there turns out to only be a total amount k of the resource available for allocation, less than the amount $\sum x_i(C)$ that they would choose to distribute if balancing only the realized cost C against potential unmet demand. This could be due to demand-side factors such as a budget cut, or supply-side factors such as manufacturing or shipping failures. In either case, we assume that these factors are undetermined at time t_0 when decisions must be made and model them with a positive random variable K giving the total amount of resource that will be available at time t_1. This adds, for each $k \in \mathbb{R}_+$, a component to our overall decision problem that asks the question: If the total amount K available of the resource is insufficient allow the procurements $x_i(C)$ you specify for the realized cost C, what alternate procurement levels $x_i^{k,C}$ would you choose? Writing $T_x = \sum_{i=1}^{N} x_i$, the action set for this new component is

$$\mathcal{X}_k = \{x \in \mathbb{R}_+^N \mid T_x = k\}. \tag{44}$$

Note that because this decision is only relevant when $\sum x_i(C) > k$ we do not include actions which use less than k total of the resource. (Why?)

In these action set variables, the loss function we used for the unconstrained N location decision problem with action set $\mathcal{X}_{cert}$ now has the expression

$$s_{TQ}(x, y) = Ck + \sum_{i=1}^{N} L(y_i - x_i)_+. \tag{45}$$

The component $x_i^{F,k}$ of the Bayes act for the contingency that $K = k$ is thus independent of the realized cost C, motivating the definition of a new single loss function (component), the *allocation loss function*

$$s_A(x, y) := L \sum_{i=1}^{N} (y_i - x_i)_+, \tag{46}$$

we use in deriving these Bayes acts components for all $\mathcal{X}_k$.

We show in section 7 that $x^{F,k}$ is defined by a single probability level τ just as was the Bayes act $x^{F,c}$ for $s_{TQ,C}$ on $\mathcal{X}_{cert}$, but that in this case, $\tau = \tau(k, F)$ depends not only on the decision problem parameter k, but also on the forecast F itself.

7   Solving the Allocation Problem

The optimal allocation $x^{F,k}$ according to a forecast F is a solution to the constrained optimization, or *allocation* problem

$$\text{(AP)} \quad \underset{x \in \mathbb{R}_+^N}{\text{minimize}} \ E_F\left[s_A(Y, x)\right] \text{ subject to } T_x = k. \tag{47}$$

The AP has a convex objective function and affine constraint, so from general convex optimization theory, a solution $x^{F,k}$ must satisfy a set of *Karush-Kuhn-Tucker* (KKT) conditions. These include that for any i such that $x_i^{F,K} > 0$,

$$\frac{\partial}{\partial x_i} E_F\left[s_A(Y, x^{F,k})\right] = L(F_i(x_i^{F,k}) - 1) = \lambda \quad \text{(cf. (9))} \tag{48}$$

for some λ independent of i. These equations follow from the assumption that when $x = x^{F,k}$, the objective $E_F[s_A(Y, x)]$ cannot be reduced by any small resource shifts between locations and they provide a quantile formula for any non-zero optimal allocation level,

$$x_i^{F,k} = F_i^{-1}(1 - \lambda/L). \tag{49}$$

Another KKT condition is the constraint equation, which can now be written in terms of λ as

$$0 = \sum_{i \in I_+} F_i^{-1}(1 - \lambda/L) - k, \tag{50}$$

where $I_+ = \{i \mid x_i^{F,k} > 0\}$. Since convexity and compactness guarantee existence of a solution, the problem is now reduced to solving (50) for the set $\{i \mid x_i^{F,k} > 0\}$ and a root $\lambda^\star$, in terms of which we can express the Bayes act as

$$x_i^{F,k} = \begin{cases} F_i^{-1}(1 - \lambda^\star/L) & i \in I_+ \\ 0 & \text{otherwise.} \end{cases} \tag{51}$$

Note, however, that this is a more complicated task than just solving (50) for $\lambda^\star$ under the assumption that $I_+ = \{1, \ldots, N\}$.[2]

The task can be simplified though by turning our attention to the *dual* formulation of the AP. (See e.g. Ruszczynski [2011], Chapter 4, which we follow here.) This involves recasting the KKT equations (48) and the constraint $T_x = k$ as the first order conditions in variables x and λ, respectively, for $(x^{F,k}, \lambda^\star)$ to be a saddle point of the *Lagrangian*

$$\mathcal{L}(x, \lambda) = \mathcal{L}(x, \lambda; F, k, L) = E_F[s_A(Y, x)] + \lambda(T_x - k) \tag{52}$$

$$= L \sum_{i=1}^N E_{F_i}[(Y_i - x_i)_+] + \lambda \left(\sum_{i=1}^N x_i - k \right). \tag{53}$$

in the region $\{x \geq 0, \lambda \in \mathbb{R}\}$.[3] The *dual function* is then defined as $\mathcal{L}_D(\lambda) := \inf_{x \in \mathbb{R}_+^N} \mathcal{L}(x, \lambda)$. From general theory, $\mathcal{L}_D(\lambda)$ is concave and $\lambda^\star$ is its unique maximizer. The reason why the dual approach is especially suitable in our situation is that our original objective function $E_F[s_A(Y, x)]$ is decomposable in the coodinates x_i, allowing the infimum $\mathcal{L}_D$ to be computed as a sum of separate infima in these coordinates:

$$\mathcal{L}_D(\lambda) = \inf_{x \in \mathbb{R}_+^N} L \sum_{i=1}^N E_{F_i}[(Y_i - x_i)_+] + \lambda \left(\sum_{i=1}^N x_i - k \right) \tag{54}$$

$$= -\lambda k + \sum_{i=1}^N \inf_{x_i \in \mathbb{R}_+} \{ L E_{F_i}[(Y_i - x_i)_+] + \lambda x_i \} \tag{55}$$

$$= -\lambda k + \sum_{i=1}^N L E_{F_i}[(Y_i - x_i(\lambda))_+] + \lambda x_i(\lambda). \tag{56}$$

Here, $x_i(\lambda)$ is any member of the set

$$\chi_i(\lambda) = \operatorname*{arginf}_{x_i \in \mathbb{R}_+} \{ L E_{F_i}[(Y_i - x_i)_+] + \lambda x_i \} \tag{57}$$

which for F_i in a standard parametric family will often simply be $\{ F_i^{-1}(1 - \lambda/L) \}$ but could an interval containing 0 or ∞. While $\mathcal{L}_D(\lambda)$ may not be differentiable when the x_i's attaining the infima in (55) are zero, it does have as a well-defined subdifferential which is the sum of intervals

$$\partial \mathcal{L}_D(\lambda) = \sum_{i=1}^N \chi_i(\lambda) - k \tag{58}$$

$$= \left[\sum_{i=1}^N \min(\chi_i(\lambda)) - k, \sum_{i=1}^N \max(\chi_i(\lambda)) - k \right] \tag{59}$$

---

[2]Such an assumption would exclude fairly simple examples in which the marginal densities f_i of F have different non-zero limits as the $x_i \to 0$.

[3]The Lagrangian in this context has the interpretation as the best (lowest) objective function value we can achieve if we forced to choose x but can buy or get credit for changes in the constraint level k at the "price" λ.

This of course looks very similar to the right hand side of (50), but the dual perspective has the key advantage of not requiring any guesses of the set $\{i \mid x_i^{F,k} > 0\}$ when initiating a search algorithm for $\lambda^\star$. Instead, given a candidate λ_0 for $\lambda^\star$, we proceed by calculating the sets $\chi_i(\lambda_0), i = 1, \ldots, N$, and using them to find the (usually zero-width) interval $\partial \mathcal{L}_D(\lambda_0)$. If $\partial \mathcal{L}_D(\lambda_0) \subset (-\infty, 0)$, we can eliminate $\{\lambda \geq \lambda_0\}$ from our search and similarly if $\partial \mathcal{L}_D(\lambda_0) \subset (0, \infty)$, we can eliminate $\{\lambda \leq \lambda_0\}$.

Intutively, we are testing to see whether the total amount $T(\lambda_0)$ of resources that would be procured at a cost $C_0 = \lambda_0$ in the unconstrained decision problem with loss function (38) leaves a deficit or surplus relative to k. T is a sum over a set of quantiles and both the set and the quantiles are non-decreasing with cost. (This is essentially why $\mathcal{L}_D$ is concave.) Thus by evaluating T at a mid-point we cut the search interval width in half at each step, arriving quickly at a close approximation $\lambda^{\star,\varepsilon}$ of the cost corresponding to k. The strategy of analysing constrained optimization problems by describing the "prices" at which unconstrained decisions match constrained ones is ubiquitous in optimization theory and its applications, especially economics.

If the F_i increase steadily enough toward 1 after leaving a neigborhood of 0, we can expect $\partial \mathcal{L}_D(\lambda)$ to be bounded away from 0 for all $\lambda \neq \lambda^\star$, allowing a binary search to converge without ambiguity. But when some F_i exhibit internal "plateaus" where their densities stay at or near 0, we can and often do in our applications encounter $\overline{\lambda}$ for which $\partial \mathcal{L}_D(\overline{\lambda})$—or at least our numerical approximation of it—contains 0 as an interior point. That is, $\overline{\lambda}$ is a price at which an uncontrained decision maker informed only by F can increase from total provision $T_1 < k$ to another $T_2 > k$ while maintaining an optimal expected utility by making (at least approximately) one-to-one trade-offs between increased resource cost and risk reduction in the one or more locations where F_i has a plateau of height near $1 - \overline{\lambda}/L$.

In this event we follow-up with a "post-processing" step that identifies lower and upper boundaries $x_{i,L}$ and $x_{i,U}$ of any such low-density plateau containing points in $F_i^{-1}(1 - \lambda^{\star,\varepsilon})$. If $F_i^{-1}(1 - \lambda^{\star,\varepsilon})$ is single-valued with the density $f_i(F_i^{-1}(1-\lambda^{\star,\varepsilon}))$ not close to zero, we simply take $x_{i,L} = F_i^{-1}(1-\lambda^{\star,\varepsilon}) = x_{i,U}$. Defining $\mathbf{x}_L = (x_{i,L})$ and $\mathbf{x}_U = (x_{i,U})$ we then interpolate an approximately optimal allocation $\mathbf{x} = (1 - t^\star)\mathbf{x}_L + t^\star \mathbf{x}_U$ where $t^\star$ solves $\sum_{i=1}^N (1 - t^\star)x_{i,L} + t^\star x_{i,U} = K$.

We note though that by using this linear interpolation (rather than some other point in the plateau satisfying the constraint) as the allocation associated with F in the scoring process we are making a choice that is not guided by any expected utility maximization principle. As such, this choice may imply strong and potentially questionable assumptions about fairness and equity principles for the decision maker when forecasts tend to be "clumped," leaving many regions with very low forecasted probability.

8 Approximating the allocation score

cdf of TY_F:

$$TF(K) = \sup \left\{ \tau \in [0,1] \mid \sum F_i^{-1}(\tau) \leq K \right\} \tag{60}$$

Given a set of distinct and increasing probability levels $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_J\} \subset (0,1)$, we can define a function

$$j_{\boldsymbol{\alpha},-}(K) = \max \left\{ j \in \{1, \ldots, J\} \mid \sum F_i^{-1}(\alpha_j) \leq K \right\} \tag{61}$$

An approximation of $S_{A,K}(F, y)$ using elicited quantiles:

$$S_{A,\boldsymbol{\alpha}}(F, y) := w_1 s_A(q_{F,j}, y) + w_2 s_A(q_{F,j+1}, y) \tag{62}$$

9 Discussion

10 References

References

A Philip Dawid. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, 59:77–93, 2007.

Tilmann Gneiting and Roopesh Ranjan. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422, 2011. doi: 10.1198/jbes.2010.08110. URL `https://doi.org/10.1198/jbes.2010.08110`.

Johannes Bracher, Evan L Ray, Tilmann Gneiting, and Nicholas G Reich. Evaluating epidemic forecasts in an interval format. *PLoS computational biology*, 17(2):e1008618, 2021.

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

A. Ruszczynski. *Nonlinear Optimization*. Princeton University Press, 2011. ISBN 9781400841059. URL `https://books.google.com/books?id=41dhYmzMSm4C`.