# Allocation scores, WIS, and CRPS via decision theory

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

August 30, 2023

**Abstract**

A somewhat open-ended attempt to motivate and construct quantile scores, WIS, CRPS, and allocation scores within a unified decision theoretic framwork.

## 1   Introduction

In this section/report/paper/rambling manifesto, 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 corresponding to the CRPS that emerge when multiple instances are confronted simultaneously. Specifically, we will consider not only whether using more or less of a 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" and the "Lagrange multiplier variables" which quantify the sensitivity of solutions 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 score a probabilistic forecast $F$ by identifying it with a solution to a *decision problem* and 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. It should be noted immediately that a better or worse score in a single instance may tell us almost nothing about the intrinsic value of the forecasting method that produced $F$. But as scores accumulate, we can use their distributions across forecasting instances to make inferences as to whether one forecasting method is better or worse than another in a decision theoretic sense.

## 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, which is determined by an action 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]

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 $s$,

---

[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.

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 $Y$.

*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 $T(F)$ (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 $T(F) = x^F$. 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. This specifies a loss function

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

where $u_+ = \max(u, 0)$. Loss functions of this general form are referred to as *piece-wise linear*, distinguishing them from smooth non-linear loss functions such as squared error. 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 over-expenditure regrets and numbers of sick individuals with unmet resource need, but the loss function $s$ maps $\mathcal{C}$ into $\mathbb{R}_+$. We also from now on make the assumption that $C < L$ since otherwise – at least according to this loss function – there is no "cost-effectiveness" in trying to prevent even a single case of unmet need. This would imply a Bayes act $x^F$ that is identically 0, telling us nothing about the value of $F$.

The Bayes act for $C \in (0, L)$ is the solution of the first order equation

$$0 = \frac{d}{dx} E_F [s_{Q,C}(x,y)] = E_F \left[ \frac{d}{dx} s_{Q,C}(x,y) \right] \tag{2}$$

$$= C + L E_F \left[ \frac{d}{dx}(y-x)_+ \right] \tag{3}$$

$$= C - L E_F [\mathbf{1}\{y > x\}] \tag{4}$$

$$= C + L(F(x) - 1), \tag{5}$$

that is, the quantile $x^F = F_l^{-1}(1 - C/L) = q_{F,\tau}$ for the probability level $\tau = 1 - C/L$. (That the critical point $q_{F,\tau}$ is actually a minimum follows from (5) 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 rule $S_Q$ on $F$ when outcome $y$ realizes by measuring the loss that occurs when the action $q_{F,\tau}$ is taken:

$$S_Q(F,y) = s_Q(q_{F,\tau}, y) = C q_{F,\tau} + L(y - q_{F,\tau})_+. \tag{6}$$

# 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 $x_i$ for a range of potential costs $C_i < L$ where $i \in 1, \ldots, m$. 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 resourse 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 $\mathcal{C} = \{C_i\}$ of possible costs by expanding the loss function to be a weighted mean of quantile losses under certainty about the cost:

$$s_{Q,\mathcal{C},p_C}(x,y) = \frac{1}{m} \sum_{i=1}^{m} p_i (C_i x_i + L(x_i - y)_+). \tag{7}$$

Here, $p_C = \{p_i\}, i = 1, \ldots, m$ is a prior distribution on $\mathcal{C}$ and $x = (x_1, \ldots, x_m)$ 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 (7) which (with some mathematical care) converts the weighted mean into an integral

$$s_{Q,f_C}(x,y) = \frac{1}{L}\int_0^L (cx(c) + L(x(c) - y)_+)f_C(c)dc. \tag{8}$$

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 varianble losses, passing to solutions of these higher dominsional 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 $(x^F(1), \ldots, x^F(m)) = (q_{F,\tau_1}, \ldots, q_{F,\tau_m})$ or a function $x(c) = q_{F,\tau(c)}$.

Similarly, given an outcome $y$, evaluation of the scoring rules $s_{Q,\mathcal{C},p_C}$ and $s_{Q,f_C}$ on the forecast $F$ amounts to vectorizing the formula for $S_Q(F,y)$. That is,

$$S_{Q,\mathcal{C},p_C}(F,y) = \frac{1}{m}\sum_{i=1}^m p_i(C_i q_{F,\tau_i} + L(q_{F,\tau_i} - y)_+) \tag{9}$$

$$S_{Q,f_C}(F,y) = \frac{1}{L}\int_0^L (cq_{F,\tau(c)} + L(q_{F,\tau(c)} - y)_+)f_C(c)dc. \tag{10}$$

How do the WIS and CRPS arise from this construction? Defining $\tau = 1 - C/L$ as a random variable gives it the density $f_\tau(\tau) = -f_C(L(1-\tau))/L$, and (10) becomes

$$S_{Q,f_C}(F,y) = \int_0^1 ((1-\tau)q_{F,\tau} + (y - q_{F,\tau})_+)f_\tau(\tau)d\tau \tag{11}$$

$$= \int_0^1 (\mathbf{1}\{y \le q_{F,\tau}\} - \tau)(q_{F,\tau} - y)f_\tau(\tau)d\tau + \int_0^1 (1-\tau)yf_\tau(\tau)d\tau \tag{12}$$

$$= w\text{CRPS}(F,y) + (1 - E[\tau])y \tag{13}$$

where $w$CRPS is the weighted CRPS of Gneiting and Ranjan [2011] with quantile weight function $\nu(\tau) = \frac{1}{2}f_\tau(\tau)$. (The $(1 - E[\tau])y$ that results from our loss function $s_Q$ adding a $Cy$ term to the standard quantile loss function does not effect the dependence of the scoring rule of $F$.)

Writing $s_{Q,\tau}(x,y) = s_{Q,C}(x,y)/L = (1-\tau)x + (y-x)_+$, we have

$$s_{Q,\tau}(q_{F,\tau},y) + s_{Q,1-\tau}(q_{F,1-\tau},y) = q_{F,\tau} + \tau(q_{F,1-\tau} - q_{F,\tau}) + (y - q_{F,\tau})_+ + (y - q_{F,1-\tau})_+ \tag{14}$$

$$= y + \tau\left(q_{F,1-\tau} - q_{F,\tau} + \frac{1}{\tau}(q_{F,\tau} - y)_+ + \frac{1}{\tau}(y - q_{F,1-\tau})_+\right) \tag{15}$$

$$= y + \tau\text{IS}_{2\tau}(F,y) \tag{16}$$

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 a means for defining the weighted interval score

$$\text{WIS}_{\alpha_{\{0:J\}}}(F, y) = \frac{1}{J + 1/2}\left(w_0|y - q_{F,.5}| + \sum_{j=1}^{J} w_j\,\text{IS}_{\alpha_j}(F, y)\right). \tag{17}$$

So by taking $\mathcal{C} = \{(1 \pm \alpha_j)L/2 \mid j > 0\} \cup \{L/2\}$ and $p_\mathcal{C}$ symmetric across $L/2$, we get

$$S_{Q,\mathcal{C},p_C}(F, y) = \text{WIS}_{\alpha_{\{0:J\}}}(F, y) + y\sum_{j=0}^{J} w_j \tag{18}$$

with $w_j = p_j\alpha_j/2$ and $w_0 = p_{L/2}$.

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

We next expand our decision problem to include multiple actions that will all, in fact (not counterfactually), have to be taken. This includes the motivating scenario for our work, the simultaneous procurement of resources across several locations (U.S. states and territories) during a pandemic. And again, summing the losses (1) over the locations (for a particular $C$ and $L$) creates a loss function that is reactive to the overall procurement decision. Let the locations be indexed by $l \in 1, \ldots, N$. The action sets for the decision problem corresponding to a single certain cost $C$, a discrete cost distribution over $\{C_i\}, i = 1, \ldots, m$, and a continuous cost distribution $f_C$ with support on $[0, L]$ are now

$$\mathcal{X}_{cert} = \mathbb{R}_+^N \tag{19}$$

$$\mathcal{X}_{disc} = (\mathbb{R}_+^m)^N = \{x(i) : \{1, \ldots, m\} \to \mathbb{R}_+^N\} \tag{20}$$

$$\mathcal{X}_{cont} = \{x(c) : [0, \infty) \to \mathbb{R}_+^N\}. \tag{21}$$

And for these "vector-action" problems we can define *total* versions of our loss functions

$$s_{TQ}(x, y) = \sum_{l=1}^{N} Cx_l + L(y_l - x_l)_+ \tag{22}$$

$$s_{TQ,\mathcal{C},p_C}(x, y) = \frac{1}{m}\sum_{l=1}^{N}\sum_{i=1}^{m} p_i(C_i x_{l,i} + L(x_{l,i} - y_l)_+) \tag{23}$$

$$s_{TQ,f_C}(x, y) = \sum_{l=1}^{N}\int_0^L (cx_l(c) + L(x_l(c) - y_l)_+)f_C(c)dc. \tag{24}$$

A key point for our presentation will be that the same $C$, set $\{C_i\}$, probabilities $\{p_i\}$, 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 $c = C_i, i \in 1, \ldots, m$) a public health worker must make in attempting to mitigate a pandemic across multiple locations $l \in 1, \ldots, N$. 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 $\sum x_l(C)$ 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 $x_l(C)$ you specify for the realized cost $C$, what alternate procurement levels $x_l^{k,C}$ would you choose? Writing $T_x = \sum_{l=1}^{N} x_l$, the action set for this new component is

$$\mathcal{X}_k = \{x \in \mathbb{R}_+^N \mid T_x = k\}. \tag{25}$$

Note that because this decision is only relevant when $\sum x_l(C) > k$ 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

$$s_{tQ}(x, y) = Ck + \sum_{l=1}^{N} L(y_l - x_l)_+. \tag{26}$$

We see then that the cost $C$ will not effect the Bayes act for this component of the decision problem motivating the definition of a new single loss function (component), the *allocation loss function*

$$s_A(x, y) := L \sum_{l=1}^{N} (y_l - x_l)_+, \tag{27}$$

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

# 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

$$(\text{AP}) \quad \underset{x \in \mathbb{R}_+^N}{\text{minimize}} \; E_F\left[s_A(Y, x)\right] \text{ subject to } T_x = k. \tag{28}$$

The AP has a convex objective function and affine constraint, so from general convex optimization theory (see e.g. Ruszczynski [2011], Chapter 4), it is equivalent to the problem of finding a saddle

point $(x^{F,k}, \lambda^\star)$ of the associated Lagrangian

$$\mathcal{L}(x, \lambda) = \mathcal{L}(x, \lambda; F, k, L) = E_F\left[s_A(Y, x)\right] + \lambda(T_x - k) \tag{29}$$

$$= L \sum_{l=1}^{N} E_{F_l}[(Y_l - x_l)_+] + \lambda\left(\sum_{l=1}^{N} x_l - k\right). \tag{30}$$

in the region $\{x \geq 0, \lambda \in \mathbb{R}\}$.[2] The Lagrangian reformulation in particular recasts the constraint $T_x = k$ of the original problem as a first-order condition on $\mathcal{L}$ with respect to the new *Lagrange multiplier* variable $\lambda$:

$$0 = \frac{\partial}{\partial \lambda} \mathcal{L}(x^{F,k}, \lambda^\star) = \sum_{l=1}^{N} x_l^{F,k} - k. \tag{31}$$

But for $(x^{F,k}, \lambda^\star)$ to be a saddle point we also need

$$0 = \frac{\partial}{\partial x_l} \mathcal{L}(x^{F.k}, \lambda) = L(F_l(x_l^{F,k}) - 1) + \lambda^\star \quad \text{(cf. (5))} \tag{32}$$

for any $x_l^{F,k} > 0$, which gives a formula

$$x_l^{F,k} = F_l^{-1}(1 - \lambda^\star/L) \tag{33}$$

for any non-zero element $x_l^{F,k}$ of the solution. Thus (31) becomes

$$0 = \sum_{\{l|x_l^{F,k}>0\}} F_l^{-1}(1 - \lambda^\star/L) - k. \tag{34}$$

Since convexity and compactness guarantee existence of a solution, the problem is now reduced to solving (31) for the set $\{l \mid x_l^{F,k} > 0\}$ and value $\lambda^\star$. Note, however, that this is a more complicated task than solving just for $\lambda^\star$ under the assumption that (33) holds for all coordinates of $x^{F,k}$.[3] The task is simplified though by bringing in the full saddle point condition for $\lambda^\star$ which implies not only (31) but also that $\lambda^\star$ is in fact the unique maximizer of the *dual function* $\mathcal{L}_D(\lambda) := \inf_{x \in \mathbb{R}_+^N} \mathcal{L}(x, \lambda)$. This approach is especially suitable in our situation because our original objective function $E_F\left[s_A(Y, x)\right]$ is decomposable in the coodinates $x_l$, allowing the dual function to be computed separately with respect to each coordinate:

$$\mathcal{L}_D(\lambda) = \inf_{x \in \mathbb{R}_+^N} L \sum_{l=1}^{N} E_{F_l}[(Y_l - x_l)_+] + \lambda\left(\sum_{l=1}^{N} x_l - k\right) \tag{35}$$

$$= -\lambda k + \sum_{l=1}^{N} \inf_{x_l \in \mathbb{R}_+} \{LE_{F_l}[(Y_l - x_l)_+] + \lambda x_l\} \tag{36}$$

$$= -\lambda k + \sum_{l=1}^{N} LE_{F_l}[(Y_l - x_l(\lambda))_+] + \lambda x_l(\lambda). \tag{37}$$

---

[2]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 changing the constraint level $k$ at the "price" $\lambda$.

[3]Such an assumption would exclude fairly simple examples in which $F$ has support bounded away from 0 in one coordinate but not in another.

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

$$\chi_l(\lambda) = \underset{x_l \in \mathbb{R}_+}{\arginf} \{LE_{F_l}[(Y_l - x_l)_+] + \lambda x_l\} \tag{38}$$

which for $F_l$ in a standard parametric family will often have simple the form $\chi_l(\lambda) = \{F_l^{-1}(1 - \lambda/L)\}$ but could an interval containing 0 or $\infty$. From general theory, $\mathcal{L}_D(\lambda)$ is concave, and while it may not be differentiable when the $x_l$'s attaining the infima in (36) are zero, it does have as a subdifferential the well-defined interval

$$\partial \mathcal{L}_D(\lambda) = \sum_{l=1}^{N} \chi_l(\lambda) - k. \tag{39}$$

This of course looks very similar to the right hand side of (34), but the dual perspective has the key advantage of not requiring any guesses of the set $\{l \mid x_l^{F,k} > 0\}$ when initiating a search algorithm for $\lambda^\star$. Instead, given a candidate $\lambda_0$ for $\lambda^\star$, we proceed by calculating the sets $\chi_l(\lambda_0), l = 1, \ldots, N$, 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 $\{\lambda \leq \lambda_0\}$.

Intuitively, 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 (22) 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 $F_l$ 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 $F_l$ 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 $F_l$ 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 Discussion

# 9 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.

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

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.

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