Evaluating infectious disease forecasts with allocation scoring rules

2023-03-30

Introduction

Infectious disease forecasts have been used as an input to public health decision-making processes such as …

In this manuscript we argue that for many decision-making problems, it is possible and valuable to develop scoring rules for probabilistic forecasts that are responsive to that decision-making context.

Literature review

Previous work in infectious disease forecasting where applications of forecasting to decision-making problems have been developed

Theoretic work on decision-theoretic set up for forecast evaluation

Previous applications of decision-theoretic evaluation to fields like economics

Previous decision-theoretic work in infectious disease forecast evaluation. I think there are at least a couple of papers out there at some level of formality. Maybe Kat Shea has something?

operations research work on constrained allocation

Methods

We first give a high-level review of a general procedure for developing proper scoring rules that are tailored to a specific decision-making task, and then use that procedure to develop a score that is suitable for evaluation of forecasts in the context of decisions about allocation of limited resources across multiple locations.

The decision-theoretic setup for forecast evaluation

In this section, we give an overview of the decision-theoretic setup for developing proper scoring rules that measure the value of a forecast as an input to decision making. We keep the discussion here at a somewhat informal level; we refer the reader to [some subset of Brehmer and Gneiting; Grünwald and Dawid; Dawid; Granger and Pesaran 2000; Granger and Machina 2006; Ehm et al. 2016] for more technically precise discussion.

In the framework of decision theory, a decision corresponds to the selection of an action x from some set of possible actions $\mathcal{X}$. For example, x may correspond to the level of investment in a measure designed to mitigate severe disease outcomes, with $\mathcal{X}$ being the set of all possible levels of investment that we might select. The quality of a decision to take a particular action x is measured in relation to an outcome y that is unknown at the time the decision is made. For example, y may correspond to the number of individuals who eventually become sick and would benefit from the mitigation measure, and informally, an action x is successful to the extent that it meets the realized demand. In the face of uncertainty, a decision-maker may use a forecast F of the random variable Y to help inform the selection of the action to take. We measure the value of a forecast as an input to this decision-making process by the quality of the decisions that it leads to.

We can formalize the preceding discussion with the following three-step procedure for developing scoring rules for probabilistic forecasts:

1. Specify a *loss function* $s(x, y)$ that measures the loss associated with taking action x when outcome y eventually occurs. We use the letter s for this function to align with the literature on forecast evaluation, in which context s may be used as a *scoring function*.
2. Given a probabilistic forecast F, determine the *Bayes act* x^F that minimizes the expected loss under the distribution F.
3. The *scoring rule* for F calculates the score as the loss incurred when the Bayes act was used: $S(F, y) = s(x^F, y)$.

This is a general procedure that may be applied in settings where it is possible to specify a quantitative loss function. Subject to certain technical conditions, scoring rules obtained from this procedure are proper (cite cite).

A review of quantile loss, CRPS, and the weighted interval score

I haven't written anything here. Fill it in, or delete this section?

The allocation score

We now develop a scoring rule for probabilistic forecasts that measures the value of a forecast as an input to decision making about how to allocate limited resources to meet demand across multiple locations. As a concrete example, we take the resource to be a good such as ventilators or oxygen supply. An administrator is tasked with determining where to send these resources so as to meet demand among hospital patients in different facilities or states.

We first specialize our notation to this decision-making setting. We define an action $\mathbf{x} = (x_1, \dots, x_n)$ as a vector specifying the amount that is allocated to each of the n locations. We require that each x_i is non-negative, and that the total allocation across all locations does not exceed a constraint K on the total available resources: $\sum_{i=1}^n x_i \leq K$. The set $\mathcal{X}$ collects all possible allocations that satisfy these constraints. The eventually realized resource demand in each location is denoted by $\mathbf{y} = (y_1, \dots, y_n)$. At the time that a decision-maker sets the resource allocation, the demand $\mathbf{y}$ is not yet known. We therefore define the random vector $Y = (Y_1, \dots, Y_n)$ where Y_i represents the as-yet-unknown level of resource demand in location i. The forecast $F = (F_1, \dots, F_n)$ collects forecasts of demand in each location. Here we identify F_i with its cumulative distribution function (CDF), and F_i^{-1} denotes the quantile function. We assume that the forecasts do not allow for the possibility of negative demand, i.e. the support of each F_i is a subset of $\mathbb{R}^+$. With this notation in place, we can proceed to develop a proper scoring rule following the outline in the previous section.

Step 1: specify a loss function. The loss associated with a particular allocation is calculated by comparing the amount allocated to each location to the realized resource demand in that location. Specifically, suppose that there is a marginal cost L that accrues for each unit of demand that is not met. We can calculate the allocation loss across all locations as

$$s(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} L(x_i - y_i)_-, \tag{1}$$

where $(x_i - y_i)_- := \max(-(x_i - y_i), 0)$ is 0 if the amount x_i allocated to unit i is greater than or equal to the realized demand y_i in that location; otherwise, it is $y_i - x_i$, the amount of unmet need in that location. We note that a number of generalizations to this loss specification have been formulated in the literature, including an allowance for costs for over-allocation to a particular unit (e.g. if there are storage costs for unused resources), differing losses different units (e.g. if a unit of unmet demand imposes more severe costs in one location than another), and the introduction of a convex function that controls the rate at which costs accrue depending on the scale of need. We consider these and other generalizations in other work, but for the present exposition we restrict our attention to the relatively simple loss formulation of Equation (1).

Step 2: Given a probabilistic forecast F, identify the Bayes act. The Bayes act is the allocation that minimizes the expected loss:

$$\mathbf{x}^F = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^N, 0\leq\mathbf{x}} \overline{s}_F(\mathbf{x}) \text{ subject to } \sum_{i=1}^N x_i \leq K, \text{ where} \tag{2}$$

$$\overline{s}_F(\mathbf{x}) = \mathbb{E}_F s(\mathbf{x}, \mathbf{Y}) = \sum_{i=1}^N \mathbb{E}_{F_i}[s(x_i, Y_i)] \tag{3}$$

It can be shown that with the loss function given in Equation (1), the Bayes act has $x_i^F = F_i^{-1}(1 - \lambda^\star/L)$, where $\lambda^\star$ is chosen so as to satisfy the equation

$$\sum_{i=1}^N F_i^{-1}(1 - \lambda^\star/L) = K. \tag{4}$$

This partial solution to the allocation problem seems to have first appeared in (Hadley and Whitin 1963); see the supplemental materials for a derivation.

One interpretation of this result is that the Bayes act sets the allocation in each location i to a quantile of the forecast distribution F_i for that location. The quantile is at a probability level $(1 - \lambda^\star/L)$ that is the same for all locations, and is chosen such that the constraint is satisfied. An alternative interpretation comes from noting that for each location i, $\frac{\partial}{\partial x_i}\overline{s}_F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^F} = \lambda^\star$ (see the supplement for a proof). In words, at the allocation given by the Bayes act, the rate of change of the expected score as a function of the amount allocated to location i is given by $\lambda^\star$. This derivative is the same for all locations, so the optimal allocation divides the available resources across all locations in such a way that according to F, the expected benefit of 1 additional unit of resources is the same in all locations.

Step 3: Define the scoring rule. We can now define a proper scoring rule for the probabilistic forecast F as

$$S(F, y) = s(\mathbf{x}^F, y) = \sum_{i=1}^n L(F_i^{-1}(1 - \lambda^\star/L) - y_i)_- \tag{5}$$

This score measures the total unmet need across all locations that results from using the Bayes allocation associated with the forecast F when the actual level of need in each location is observed to be y_i.

ELR: This section would likely benefit from a figure or two to illustrate the ideas...?

Description of what happens if there is uncertainty about the the constraint K. Integration across K results in something like a forecaster-specific weighted CRPS. It might help to introduce notation like $\lambda^\star(K)$ or $S_K(F, y)$ throughout all of the above discussion? Then if we have a prior $p(K)$ on K, we get to something like

$$S(F, y) = \int S_K(F, y)p(K)\, dK$$

as our final score.

Application

We illustrate with an application to hospital admissions in the U.S., considering the problem of allocation of a limited supply of medical resources to the states.

Case study heading into the Omicron wave. Some more detailed discussion of implications of bad forecasts for specific decision-making purposes – take a "deep dive" into one or two example states like FL.

Look at results over a broader range of time.

Discussion

We often conceive of infectious disease forecasts as being useful for decision-making purposes, but it is rare for forecast evaluation to be tied directly to the value of the forecasts for informing those decisions. This work seeks to address that gap.

We have demonstrated that evaluation methods that are tied to decision-making context can yield model rankings that are substantively different from generic measures of forecast skill like WIS.

In practice, there are many users of forecasts with many different decision-making problems. Not all can be easily quantified. Those that can be easily quantified may differ enough that no single score is appropriate for all users. We suggest reporting multiple scores. This may be tricky to operationalize in the setting of a general forecast hub. It matters how you elicit and represent probabilistic forecasts (quantiles? samples? cdfs?).

The allocation score we developed here does not directly account for important considerations such as fairness/equity of allocations.

The allocation score we developed also does not attempt to capture the broader context of decision-making. For example, in practice it may be possible to increase the resource constraint K by shifting funding from other disease mitigation measures.

Forecaster's dilemma: a successful forecast may lead to decisions that change the distribution of the outcome Y. Our framework cannot be used in those settings.

There is much more to do in this general area.

References

Hadley, G., and Thomson M. Whitin. 1963. *Analysis of Inventory Systems.* Prentice-Hall International Series in Management. Prentice-Hall.