Comments on set up for application

2023-05-22

Notation:

- s is the date a forecast is made
- t is the target date for a forecast
- $y_t = (y_{1,t}, \dots, y_{l,t}, \dots, y_{L,t})$ is the vector of realized resource need in each of L locations on the target date t
- K_t is the constraint on available resources that is in effect for target date t. For example, this may change over time if the planning agency is going to ramp up (or down) production of the resource in question over time.
- $x_t = (x_{1,t}, \dots, x_{l,t}, \dots, x_{L,t})$ is the vector of resource allocation levels for target date t, with $\sum_l x_{l,t} \leq K_t$.

Preliminary comments:

- The introduction of the time indices s and t is new in this document. Everything in our main manuscript is framed for a single forecast (so, one value of s) for a single target date t. But our application now considers multiple forecast dates and target end dates, so we need this notation to think carefully about what's going on.
- Our goal is to figure out how to apply the ideas we've already developed to multiple forecast dates in a justifiable way. Other work (e.g. Bertsimas et al.) has more carefully considered how to set up the allocation problem in a way that incorporates planning over time, but we don't want to go there for this paper.

Concerns, high level:

1. Currently, a working decision was made in the application to take "the alloscore value that was for the K closest to the actual total observations," which I'll denote here as $y_t^{tot} = \sum_l y_{l,t}$. I think there are two problems with this:
   (a) It's not clear to me whether or not this choice for K_t likely yields an improper score. The basic reason is that the score depends on the value of the random variable/vector that is being forecasted. This may or may not be OK?
   (b) It's hard to see how to justify this choice for K_t in a way that corresponds to the set up of the decision-making problem. The implication is that at each time t, we have exactly enough resources to meet what will be the eventually-realized need (if an oracle forecast is used). But this is not realistic.
2. Closely related to 1 (b): How are we thinking about the set up of the decision making problem across multiple dates? Is there a way we could summarize overall model performance across the time span under consideration with a single number? To the extent possible, I think we should try to start from a plausible multi-time-step decision making set up that can fit within our set up and go from there.
   (a) As described in 1.b., I don't think setting $K_t = y_t^{tot} = \sum_l y_{l,t}$ can be justified as corresponding to any natural decision making set up. In general, K_t will not be equal to y_t^{tot}; our resource planning is not that good/lucky.

(b) If I remember right, on our call a few weeks ago Ben suggested something like "Suppose you get one chance to allocate resources with a constraint K, and then those allocation levels are fixed and carry through the wave under consideration." This seems like a justifiable idea, but it doesn't line up well with the kind of analysis we think we'd like to do in our application where we consider multiple forecast dates s. Looking at multiple forecast dates only makes sense if we can update our allocations based on the information in those forecasts.

(c) Another idea is to say that we have K units of resources and on each forecast date s, we get to decide how they will be allocated on target date $t = s + h$.

- This is a little dicey if the resource is something fixed like ventilators or hospital beds because some of the resources we had previously allocated might still be in use by the time t rolls around (e.g. if someone is hospitalized or on a ventilator for multiple weeks)
- Could make more sense for something like supply of oxygen or medication, if K new units of the resource are manufactured each week.
- Recognize that maybe we don't want to talk too much about how this connects to real life in the manuscript...
- But I don't have any great ideas for how we might motivate any particular fixed value of K of interest ahead of time without trying to connect to context (e.g. look up some information somewhere about total oxygen supplies and average amounts used per admitted hospital patient...)
- The strategy I took when talking through things in london was to try not to fixate too much on any particular value of K, but to note that model rankings were fairly stable across values of K.

(d) Returning to the question I raised above about how we might summarize overall model performance across the time span under consideration: it seems like maybe if we pre-specify a grid of values of K that are used for all target dates t, for each of those values of K maybe we could average scores across the target dates? I think this is basically Ben's suggestion, but without the idea that we only get to specify an allocation once. We'd maybe end up with something like the peaked plot, but with multiple peaks, one for each target date...? If you had enough target dates, maybe you'd get to something relatively smooth/domed. This is still not a single number, and there is still the question of how to summarize across K's to get to a single number.

More on propriety

Some general set up that's somewhere between Gerding et al. and Gneiting and Raftery:

- $(\mathcal{Y}, \mathcal{A})$ is a measurable space
- $\mathcal{P}$ is a convex class of probability measures on $(\mathcal{Y}, \mathcal{A})$
- A probabilistic forecast is any probability measure $F \in \mathcal{P}$
- $S(F, y)$ is a scoring rule
- $\overline{S}(F, G) = \mathbb{E}_{Y \sim G}[S(F, Y)]$ is the expected score for F under the distribution G
- The (negatively oriented) scoring rule S is proper relative to $\mathcal{P}$ if $\overline{S}(G, G) \leq \overline{S}(F, G)$ whenever $F, G \in \mathcal{P}$
- $s(x, y)$ is a loss function [or scoring function if x is a functional of F]
- We can obtain a scoring rule from a scoring function by setting $S(F, y) = s(x^F, y)$ where $x^F = \text{argmin}_x \mathbb{E}_{Y \sim F}[s(x, Y)]$

Specifics for our set up for allocation:

- Our scoring function is $s_K(x,y) = \sum_{i=1}^n L(y_i - x_i)_+$ (subject to constraints that each x_i is non-negative and $\sum_i x_i \leq K$, which I suppose we could build into s_K)
- x^F has entries $x_i^F = F_i^{-1}(1 - \lambda^{\star,F}(K)/L)$ where $\lambda^{\star,F}(K)$ is chosen such that $\sum_i x_i^F = K$
- Our scoring rule is $S_K(F,y) = s_K(x^F, y) = \sum_{i=1}^n L(y_i - F_i^{-1}(1 - \lambda^{\star,F}(K)/L))_+$

Since S_K is derived from a scoring rule using Bayes actions, it is a proper scoring rule. Writing this out a bit, we have to check that for all $F, G \in \mathcal{P}$:

$$
\begin{aligned}
& \overline{S}(G,G) \leq \overline{S}(F,G) \\
\Leftrightarrow\quad & \mathbb{E}_{Y\sim G}\left[S_K(G,Y)\right] \leq \mathbb{E}_{Y\sim G}\left[S_K(F,Y)\right] \\
\Leftrightarrow\quad & \mathbb{E}_{Y\sim G}\left[\sum_{i=1}^n L(Y_i - G_i^{-1}(1 - \lambda^{\star,G}(K)/L))_+\right] \leq \mathbb{E}_{Y\sim G}\left[\sum_{i=1}^n L(Y_i - F_i^{-1}(1 - \lambda^{\star,F}(K)/L))_+\right].
\end{aligned}
$$

But for any given K this inequality holds by construction, because we chose the Bayes act allocation by minimizing the left hand side.

What about when $K = Y^{tot} = \sum_i Y_i$? Continuing from the last line above, we have

$$
\mathbb{E}_{Y\sim G}\left[\sum_{i=1}^n L(Y_i - G_i^{-1}(1 - \lambda^{\star,G}(Y^{tot})/L))_+\right] \leq \mathbb{E}_{Y\sim G}\left[\sum_{i=1}^n L(Y_i - F_i^{-1}(1 - \lambda^{\star,F}(Y^{tot})/L))_+\right]
$$

I can't decide:

- Is the result now immediate through some kind of iterated expectations thing?
- Or is this now complicated because Y^{tot} (which is a function of all of the Y_i) appears in a nonlinear way in each of the terms of the summation, so that we can no longer separate into terms by location?