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

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

November 13, 2023

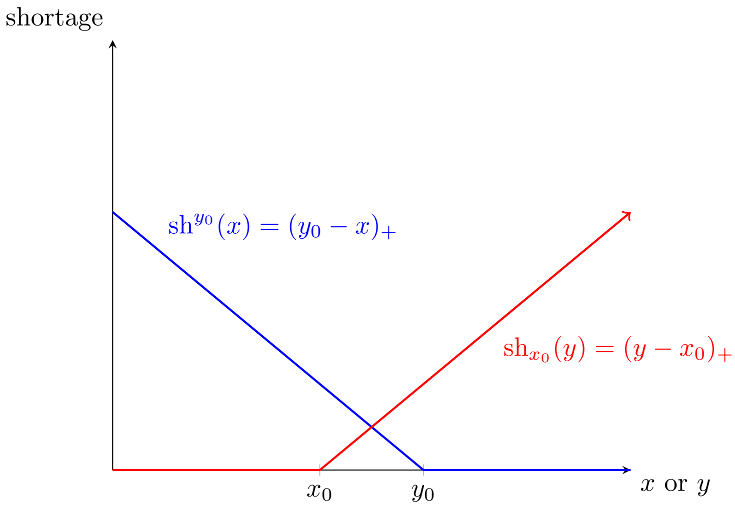
Abstract

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?
- Explain "All proper scoring rules for probabilistic forecasts have an explicit link to a loss function" from discussion.
- 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 exponetial 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.
- 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.
- Sketch of scoring for decision problems involving both cost and constraint.

1 Shortages

For convenience, we write $u_+ = \max\{0, u\}$, and refer to $(y - x)_+$ as a *shortage* in accordance with our typical use of y for a demand or need and x for an available supply. To regard shortage as a function depending on 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:



Let Y be a random variable with distribution F. 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. We see then that the *expected shortage*[1] $\mathbb{E}_F[(Y - x)_+] = \mathbb{E}_F[\mathrm{sh}^Y](x)$ (assuming it exists) is also convex (and therefore continuous) in x by integrating the convexity inequality for $\mathrm{sh}^y(x)$ with respect to the probability measure $dF(y)$:

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

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

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

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 shows that if F does not

[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 *conditonal* expectation $\mathbb{E}_F[Y - x \mid Y - x \geq 0] = \mathbb{E}_F[(Y - x)_+]/\mathbb{P}_F\{Y \geq x\}$.

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

$$\frac{d}{dx}\,\mathbb{E}_F[(Y-x)_+] = \frac{d}{dx}\int_x^\infty (y-x) f_Y(y)dy \tag{12}$$

$$= \int_x^\infty \frac{d}{dx}(y-x) f_Y(y)dy - (x-x)f_Y(x) = -\int_x^\infty f_Y(y)dy = F(x) - 1. \tag{13}$$

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

2 Quantiles and Expected Shortage

We recall how quantiles arise as solutions to a probabilistic decision problem. Let Y be a random variable representing the future level of an undesirable outcome such as severe COVID incidence. Let x be a decision variable representing the possible levels of some costly counter-measure, such as procurement of monoclonal antibody treatments, that can be taken at a cost C per unit in preparation for Y. 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 economic value) and combine the total shortage loss with expenditure into a *loss function*[2]

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

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[3] 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, 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

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

[2] 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.

[3] 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.

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. Thus $x^\star$ minimizes the expected loss (14) and gives an optimal sultion to the decision problem if and only if it is a quantile $q_{\alpha,F}$.

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

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

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

3 Allocation Bayes acts as vectors of marginal quantiles.

Here we explain why the Bayes act $x_i^{F,K}$ for the allocation problem (4) 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, $x_i^{F,K} = F_i^{-1}(\tau^{F,K})$.

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

$$\mathbb{E}_F[s_A(x, Y)] = \sum_{i=1}^{N} L \cdot \mathbb{E}_{F_i}[(Y_i - x_i)_+] \tag{21}$$

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,

$$\mathbb{E}_{F_i}[(Y_i - x_i^\star - \delta)_+] - \mathbb{E}_{F_i}[(Y_i - x_i^\star)_+] \geq \mathbb{E}_{F_j}[(Y_j - x_j^\star)_+] - \mathbb{E}_{F_j}[(Y_j - x_j^\star + \delta)_+]. \tag{22}$$

Since the expected shortages in i and j have right and left derivatives at any x_i and x_j (see Section 1), we can divide (22) by δ and take limits for $\delta \searrow 0$ to get

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

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), (23) becomes

$$L(1 - F_i(x_i^\star-)) \geq L(1 - F_j(x_j^\star)). \tag{24}$$

Inequalities (23) and (24) 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 single number λ (a *Lagrange multiplier*) exists such that

$$L(1 - F_i(x_i^\star-)) \geq \lambda \geq L(1 - F_i(x_i^\star)), \quad \text{for all } i \in 1, \ldots, N \tag{25}$$

that is,

$$F_i(x_i^\star) \geq 1 - \lambda/L \geq F_i(x_i^\star-), \tag{26}$$

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

4 Properties and Properness

For a prediction to be useful, it must **proper**ly describe a **proper**ty.