1 Definitions

We model a forecasting competition as m independent events denoted by the independent binary random variables $X_k \in \{0, 1\}$ for $1 \le k \le m$. Specifically, let $\theta_k = P(X_k = 1) = 1 - P(X_k = 0)$. For each event, we draw the outcome for the event $x_k \sim X_k$. We say event k occurs when and only when $x_i = 1$. There are n participants in the forecasting competition. Each forecaster independently submits a vector of predictions $y_i \in [0, 1]^m$, such that $y_{i,k}$ is forecaster i's estimate of θ_k. We call these y_i the forecasts of the ith forecaster.

A forecasting mechanism $M(y, x)$ maps the forecasts $y = (y_1, ..., y_n)$ and outcomes of the events $x = (x_1, ..., x_k)$ to an index $w \in [1, n]$. We say that forecaster w wins the competition if $M(y, x) = w$. Next, we consider some properties that an ideal mechanism would have.

We call a mechanism truthful iff, for all forecasters, their probability of winning is maximized when they reports their beliefs as their forecasts. Specifically, let $\hat{\theta}_{i,k}$ be forecaster i's belief of the value θ_k. We say that M is truthful if for any $y_i \ne \hat{\theta}_i$,

$$P(M((y_1, ..., \hat{\theta}_i, .., y_n), x) = i) \ge P(M((y_1, ..., \hat{y}_i, .., y_n), x) = i)$$

and we say M is strictly truthful if the strict inequality holds.

Additionally, we would like our mechanism to choose the forecaster with the best predictions. We define the accuracy of a single forecaster i as 1 minus the average squared loss of their forecasts compared to the true probabilities. Specifically, $a_i = 1 - \frac{1}{m} \sum_{k=1}^{m} (y_{i,k} - \theta_k)^2$. An ideal mechanism would choose forecasters with larger accuracies with higher probabilities. Let $i^* = \arg\max_i a_i$ be the highest accuracy forecaster, and let $a^* = a_{i^*}$. We call $\epsilon = a^* - \max_{i \ne i^*} a_i$ the accuracy gap. ϵ is the difference in accuracy between the best and second best forecaster.

For a given mechanism M and n forecasters with accuracy gap ϵ, we let m_δ be the minimum number of events such that M necessarily chooses i^* with probability at least $1-\delta$. Specifically, for any $m > m_\delta(M, n, \epsilon)$, $P(M(y, x) \ne i^*) \le \delta$. m_δ gives us a measure of how many events are needed for a mechanism to stably pick the best forecaster. We will use it as a metric to compare mechanisms, where we say M is more efficient than M' if $m_\delta(M, n, \epsilon) < m_\delta(M', n, \epsilon)$, since M needs less events to consistently pick the best forecaster.

define proper scoring rule Let p denote belief and y denote report A scoring rule R is proper if $\forall p, y \in 0, 1, E_{x\ p}[R(p, x)] \ge E_{x\ p}[R(y, x)]$

2 Simple average mechanism

The simplest selection mechanism (and what is often done in practice), we call the simple average mechanism. For this mechanism, we assign each forecaster a score $f_{i,k} = R_q(y_{i,k}, x_k)$ for each event. Then, we assign their final score by summing these over all events, $F_i = \sum_{j=1}^{m} f_{i,j}$. Finally, we choose the forecaster with the highest cumulative score $\arg\max_i F_i$. Each forecaster's cumulative score in this mechanism $\sum_{i=1}^{m} 1 - (y_{i,j} - x_i)^2$ is an analog for their cumulative accuracy, $ma_j = \sum_{i=1}^{m} 1 - (y_{i,j} - \theta_j)^2$. However, we will show that this approximation can be flawed.

2.1 Truthfulness

show simple counterexample(s?)
- single event favors extremes
- multiple events can be dominated by single untruthful event when rest are isomorphic

2.2 Accuracy

Assume forecasters remain honest. For single events, the most extreme forecasters will get selected.

For multiple events, we will see that we get similar accuracy bounds as the SHELF mechanism, assuming the forecasters remain honest. In fact, the main advantage of SHELF is that it is as efficient as the Naive mechanism, but also elicits truthful forecasts.

3 ELF

The ELF mechanism, introduced in/by [cite elf paper], improves upon the Naive mechanism, by incentivizing forecasters to be truthful. For each event k, we hold a lottery to choose a winner. Forecaster i is chosen with probability

$$f_{i,k} = \frac{1}{n} + \frac{1}{n}\left(R_q(y_{i,k}, x_k) - \frac{\sum_{j \neq i} R_q(y_{j,k}, x_k)}{n-1}\right)$$

This score is the additively normalized quadratic score among all forecasters. A forecaster increasing their own quadratic score increases the chances they win an event, while decreasing the chances others win that event. Let w_k be the winner of event k. The ELF mechanism, M_l, then chooses $\arg\max_i \sum_{k=1}^{m} \mathbb{1}(w_k = i)$ as the winner of the competition. Essentially, we give the winner of each event a point, and then pick the forecaster who got the most points.

In [elf paper] Theorem 6, they show that ELF is strictly truthful. In Theorem 8, they also show it is accurate. Specifically, that for an accuracy gap of ϵ, ELF has

$$m_\delta \leq \frac{2(n-1)^2}{\epsilon^2} \ln\left(\frac{4(n-1)}{\delta}\right)$$

For a fixed forecaster accuracy gap and desired mechanism accuracy, the number of events ELF needs is proportional to $n^2 \ln n$. However, this upper bound is not tight, and can be lowered significantly.

Let $F_{i,k}$ be the indicator function for forecaster i winning event k. Then, $F_i = \sum_k F_{i,k}$ is the random variable that is the number of events forecaster i wins. The proof of [elf paper] Theorem 8 utilized the Hoeffding bound applied to F_i, which assumes they have high variance. However, this is not the case, which leads to an improvement in the lower bound.

Theorem 1 (Bernstein's Inequality) *Given random variables X_i for $1 \leq i \leq n$ such that $0 \leq X_i \leq 1$ almost surely, let $S = \sum_i X_i$. Then,*

$$P\left(|S - \mathbb{E}[S]| < t\right) < 2\, e^{\frac{-t^2}{2\left(\mathrm{Var}(S) + \frac{t}{3}\right)}}$$

Theorem 2 *For an accuracy gap ϵ, ELF correctly chooses the most accurate forecaster with probability at least $1 - 4(n-1)e^{-\frac{m\epsilon}{6(n-1)}}$.*

Proof: Note that for any single event k, every forecaster wins with probability at best $\frac{2}{n}$, and at worst $\frac{1}{n} - \frac{1}{n(n-1)}$ (achieved when $y_i = 1$, $y_{j \neq i} = 0$, and $x_i = 1$). Since $F_{i,k}$ is just a binary random variable, it will have highest variance when it's expectation is as close to $\frac{1}{2}$ as possible. For nontrivially small n, this will mean its expectation is $\frac{2}{n}$ (or equivalently $P(F_{i,k} = 1) = \frac{2}{n}$), so its variance will be at most $\frac{2(n-2)}{n^2}$. So we have $\mathrm{Var}(F_{i,k}) \leq \frac{2(n-2)}{n^2} < \frac{2}{n}$.

Since $F_i = \sum_k F_{i,k}$, we have $\mathrm{Var}(F_i) = \sum_k \mathrm{Var}(F_{i,k}) < \sum_k \frac{2}{n} = \frac{2m}{n}$. As shown in [elf paper], for $j \neq i$ we have

$$\mathbb{E}[F_i] - \mathbb{E}[F_j] \geq \frac{m\epsilon}{n-1}$$

Therefore, if $F_j \geq F_i$, (forecaster j beats forecaster i), then either forecaster j overperformed their expectation by at least $\frac{m\epsilon}{2(n-1)}$, or forecaster i underperformed their expectation by $\frac{m\epsilon}{n-1}$. Specifically, either $\mathbb{E}[F_i] - F_i \geq \frac{m\epsilon}{2(n-1)}$ or $F_j - \mathbb{E}[F_j] \geq \frac{m\epsilon}{2(n-1)}$

Applying Theorem 1, we have

$$
P\left(|F_i - \mathbb{E}[F_i]| < \frac{m\epsilon}{2(n-1)}\right) < 2\, e^{\frac{-\left(\frac{m\epsilon}{2(n-1)}\right)^2}{2\left(\frac{2m}{n} + \frac{m\epsilon}{6(n-1)}\right)}}
$$

$$
< 2\, e^{\frac{-\left(\frac{m\epsilon}{2(n-1)}\right)^2}{2\left(\frac{m}{n-1}\left(2 + \frac{\epsilon}{6}\right)\right)}}
$$

$$
= 2\, e^{-\frac{\frac{m\epsilon^2}{2(n-1)}}{(2 + \frac{\epsilon}{6})}}
$$

$$
< 2\, e^{-\frac{m\epsilon^2}{5(n-1)}} \tag{1}
$$

And similarly for F_j. Putting both of those cases together, we have by the union bound

$$
P(F_j \geq F_i) = P\left(\left(F_j - \mathbb{E}[F_j] < \frac{m\epsilon^2}{2(n-1)}\right) \cup \left(\mathbb{E}[F_i] - F_i < \frac{m\epsilon^2}{2(n-1)}\right)\right)
$$

$$
\leq P\left(F_j - \mathbb{E}[F_j] < \frac{m\epsilon^2}{2(n-1)}\right) + P\left(\mathbb{E}[F_i] - F_i < \frac{m\epsilon^2}{2(n-1)}\right)
$$

$$
\leq 4\, e^{-\frac{m\epsilon^2}{5(n-1)}}
$$

Using the union bound over all j, we have

$$
P\left(M_l(y_1, ..., y_n, x) = i\right) = 1 - \sum_{j \neq i} P\left(M_l(y_1, ..., y_n, x) = j\right)
$$

$$
\geq 1 - \sum_{j \neq i} P\left(F_j \geq F_i\right)
$$

$$
\geq 1 - 4(n-1)e^{-\frac{m\epsilon^2}{5(n-1)}}
$$

$\square$

Assigning this probability to $1 - \delta$, we see that for a fixed n, ϵ, ELF will choose the correct forecaster with probability at least $1 - \delta$ if m is large enough for the above inequality to hold. Solving it for m, we get the minimum such satisfying value

$$
m_\delta \leq \frac{5(n-1)}{\epsilon^2} \ln\left(\frac{4(n-1)}{\delta}\right)
$$

4

4 SHELF

For this mechanism, we let the binary random variable $X_{i,k}$ be 1 with probability $R_q(y_{i,k}, x_i)$, and 0 otherwise. Then, for each event k we assign forecaster i the score $f_{i,k} = X_{i,k}$ obtained by sampling this variable. Finally, we assign their final score by summing these over all events, $F_i = \sum_{j=1}^{m} f_{i,j}$, then chose the forecaster with the highest F_i, breaking ties uniformly at random.

4.1 Truthfulness

(following section 4.2, theorem 4) For a single event with true probability θ, every forecaster will either win 1 or 0 events. The probability that their final score is 1 will just be their expected score for the single event. Therefore, to maximize their final score, each forecaster must maximize their expected score for the single event. Their score is just $\mathbb{E}_{x\sim\theta}(R_q(y_i, x)) = \mathbb{E}_{x\sim\theta}(1 - (y_i - x)^2)$. This is maximized when $y_i = \theta$, so the mechanism is truthful. (what theorem 4 does) We can also see this directly be noting that R_q is a strictly proper scoring rule, so its expectation is maximized when $y_i = \theta$.

Theorem: SHELF is truthful for $m \geq 1$ events.

Proof. For each forecaster $j \in [1, n]$ fix forcaster's report.

S'pose that for some forecaster i, i's report on some subset of events, where $k \in [1, m]$ denotes the size of this subset, does not maximize their expected score for any of those individual events.
Let k' be one of those events.

We will now show that i will maximize their probability of being chosen as the best forecaster in the SHELF mechanism iff they alter their k reports to maximize their expected score on each k events. A guiding intuition here would be as follows: The higher the expected score a forecaster has on a given event, the higher the probability that SHELF awards them a point on this specific event, and the only way to maximize one's probability of being selected as the best forecaster is to increase the number of points they are awarded. Therefore, for a single independent event, increasing one's expected score will either increase the probability of being selected as the best forecaster or leave it unchanged, and it does so without altering results of any other event. In this way, we will show that being truthful for a singular event will always increase one's proba-

bility of being selected as the best forecaster. We can then repeat this process for each untruthful report to maximize the probability of being selected as the best forecaster.

Formally, we will analyze three cases relating to the context provided by the $m - 1$ events other than k'.

Let $V(i)$ denote the number of points forecaster i has accrued.
Let $f_{i,k}$ denote the probability forecaster i scores a point on event k.

1. Some forecaster j scores at least two more points than i, OR i scores at least two more points than any other forecaster.

$$\exists j \in [1, m] \text{ s.t. } j \neq i \text{ and } V(j) - V(i) \geq 2$$

$$OR$$

$$\forall j \in [1, m] \text{ s.t. } j \neq i; V(i) - V(j) \geq 2$$

For this case, i's probability of being selected as the best forecaster is either 1 or 0. It is also unaffected by being truthful or not in k', because gaining, or failing to gain, 1 point in this context is not enough to alter the outcome as only the forecaster with the most number of points is chosen as the best forecaster with ties broken uniformly.

2. Some forecaster j scores exactly one more point than i, AND no other forecasters score strictly more points than i.

$$\exists j \in [1, m] \text{ s.t. } j \neq i \text{ and } V(j) - V(i) = 1$$

$$AND$$

$$\forall l \in [1, m] \text{ s.t. } l \neq i \text{ and } l \neq j; V(i) - V(l) \geq 0$$

For this case, i must seek to maximize $f_{i,k'}$ to have any non-zero probability of being selected as the best forecaster under the SHELF mechanism. In the SHELF mechanism, the maximum payout a forecaster can receive for a single event is 1, and that point is awarded to them with probability exactly equal to their quadratic score. Thus, we know a forecaster's expected value for a singular event is exactly their quadratic score. Since the quadratic scoring rule is proper, we know that the expectation of a

6

forecaster's quadratic score is maximized when the forecaster is reporting their true beliefs. As a result, we can conclude that the only way for i to maximize their probability of receiving a point in k' is to report their true beliefs. By reporting their true beliefs, which maximizes the probability that they score another point, i also maximizes the probability of being selected as the best forecaster which, in this context, occurs when only i scores a point and the tie between j and i is broken uniformly via the rules of SHELF.

3. No forecasters score more points than i, but there is at least one forecaster j that scores the same number of points as i or scores exactly one less point than y_i.

$$\exists j \in [1, m] \text{ s.t. } j \neq i \text{ and } V(i) - V(j) \in \{0, 1\}$$

In any given configuration, a forecaster's probability of being selected as the best forecaster is maximized when the minimal number of other forecaster have as many points as them. In the SHELF mechanism, any forecaster's probability of scoring a point on a given event is independent of all other. Therefore, there is no action that can be taken to minimize the probability of any competitor's probability of being selected as the best forecaster except strictly increasing one's own point total. In this final case, the only way for i to increase their probability of being chosen as the best forecaster would be to increase the probability of scoring a point in k'. The same logic as case 2 holds here that i would then prefer to be truthful and maximize their expected score in k' to increase their probability of being selected as the best forecaster.

Since any configuration of the $m-1$ events falls into one of these three cases, we have shown that any forecaster is weakly incentivized to report truthfully. To show that any forecaster is strongly incentivized to report truthfully, we simply show that i believes cases 2 or 3 will happen with positive probability. Since $y_{j,k} = p_{j,k} \in (0, 1) \forall j \forall k$, we know that every forecaster has some uncertainty on each event. As a result, we can conclude that $f_{j,k} > 0 \forall j \forall k$. We can think of all possible outcomes as the set of all possible binary strings of length n, and since $f_{j,k} > 0 \forall j \forall k$ there exists at least one sequence of outcomes strings picked from our set that constitutes cases 2 or 3. As a result, there is a positive probability that cases 2 or 3 occur, and thus forecasters are strongly incentivized to report truthfully.

4.2 Accuracy

From proposition 2, we have that $E_{x \sim \theta}(R_q(y, x)) = R_q(y, \theta) + \theta^2 - \theta$. Additionally, for a single event, a forecaster's accuracy a_i is by definition just $R_q(y, \theta)$. So, $E_{x \sim \theta}(R_q(y, x)) = a_i + \theta^2 - \theta$. For a single event with probability θ, forecaster i's expected score is $\mathbb{E}_{x \sim \theta} f_i = E_{x \sim \theta}(R_q(y_i, x)) = a_i + \theta^2 - \theta$.

4.3 Truthfulness for a single event

(section 4.3, proposition 5). For forecaster i to be selected, they either need to have $f_i = 1$, and then be chosen uniformly at random from all others who scored 1, or everyone must score 0 and they will be chosen uniformly at random. In the first case, the expected number of other forecasters who also score 1 is $\sum_{i' \neq i} \mathbb{E}_{x \sim \theta} f_{i'} = \sum_{i' \neq i}(a_{i'} + \theta^2 - \theta) = (n-1)(\theta^2 - \theta) \sum_{i' \neq i} a_{i'}$. Then, the probability that forecaster i is chosen is

$$P(f_i = 1)\frac{1}{1 + (n-1)(\theta^2 - \theta)\sum_{i' \neq i} a_{i'}} = \frac{a_i + \theta^2 - \theta}{1 + (n-1)(\theta^2 - \theta)\sum_{i' \neq i} a_{i'}}$$

In the other case, the probability i wins is just

$$\frac{1}{n} \prod_i P(f_i = 0) = \frac{\prod_i (1 - a_i - \theta^2 + \theta)}{n}$$

There's probably a better way to do this... Actually coming back to this I'm not even sure that the first step is allowed (using expected number of winners directly instead of summing over cases)

4.4 Truthfulness for multiple events

multiple events: (section 5.3, theorem 8) Let $a_i \geq a_j + \epsilon$ for all $j \neq i$. Like in theorem 8, we first bound the expected number of events won that i exceeds

8

j by.

$$\mathbb{E}\left(\sum_{k=1}^{m} f_{i,k}\right) - \mathbb{E}\left(\sum_{k=1}^{m} f_{j,k}\right) = \left(\sum_{k=1}^{m} \mathbb{E}f_{i,k}\right) - \left(\sum_{k=1}^{m} \mathbb{E}f_{j,k}\right)$$

$$= \sum_{k=1}^{m} \left(\mathbb{E}(f_{i,k}) - \mathbb{E}(f_{j,k})\right)$$

$$= \sum_{k=1}^{m} \left((a_i + \theta_k^2 - \theta_k) - (a_j + \theta_k^2 - \theta_k)\right)$$

$$= \sum_{k=1}^{m} (a_i - a_j)$$

$$\geq \sum_{k=1}^{m} \epsilon$$

$$= m\epsilon$$

If $F_j \geq F_i$, then either $\mathbb{E}(F_i) - F_i \geq \frac{m\epsilon}{2}$ or $F_j - \mathbb{E}(F_j) \geq \frac{m\epsilon}{2}$. By Hoeffding's, we can upper bound the probability of each of these events by $2e^{-\frac{2(\frac{m\epsilon}{2})^2}{m}} = 2e^{\frac{-m\epsilon^2}{2}}$. Using the union bound, the probability that either of those events occurs is twice this, or $4e^{\frac{-m\epsilon^2}{2}}$.

Using the union bound again over all $j \neq i$, we have

$$1 - \delta = P(M(y_1, ..., y_n, x) = i) = 1 - \sum_{j \neq i} P(F_j \geq F_i) = 1 - 4(n-1)e^{\frac{-m\epsilon^2}{2}}$$

Letting this quantity be $1 - \delta$ and solving for m, we see that this mechanism correctly chooses forecaster i with probability $1 - \delta$ when m is large enough to satisfy the inequality. Solving for m gives us the minimum such value, which is an upper bound of m_δ for SHELF.

$$m_\delta \leq \frac{2}{\epsilon^2} \ln \frac{4(n-1)}{\delta}$$