Forecasting Competitions with Correlated Events

Rafael Frongillo, Manuel Lladser, Anish Thilagar, Bo Waggoner

University of Colorado Boulder

Abstract

Beginning with Witkowski et al. (2023), recent work on forecasting competitions has addressed incentive problems with the common winner-take-all mechanism. Frongillo et al. (2021) propose a competition mechanism based on Multiplicative Weights, an online learning algorithm. They show that their mechanism selects an ϵ-optimal forecaster with high probability using only $O(\log(n)/\epsilon^2)$ events. These works, together with all prior work on this problem thus far, assume that events are independent. We prove the first accuracy and approximate truthfulness guarantees for forecasting competitions with correlated events. To quantify correlation, we introduce a notion of block correlation, which allows each event to be strongly correlated with up to b others and weakly correlated with the rest. We show that under distributions with this correlation, the Multiplicative Weights mechanism retains its ϵ-optimal guarantee using $O(b^2 \log(n)/\epsilon^2)$ events. Our proof involves a novel concentration bound for correlated random variables which may be of broader interest.

Extended version — https://arxiv.org/abs/2303.13793

1 Introduction

Forecasting competitions, such as those on Kaggle or the Good Judgement project, attempt to discern which forecaster from a pool of contestants has the best forecasting model. Traditionally, competition mechanisms ask each of n forecasters to predict the probability of m future events, and upon seeing the event outcomes, pick the forecaster with the largest empirical score as the winner. As many have noted, this approach distorts the incentives of the forecasters, who will extremize their reports in order to increase their chances of having the maximum empirical score [Lichtendahl and Winkler 2007, Kaggle 2017, Witkowski et al. 2018, Aldous 2019, Frongillo et al. 2021, Witkowski et al. 2023]. When their reports deviate from their models' true predictions, it becomes unclear which forecasters are actually the best at forecasting, as opposed to being better at strategizing, leaving no guarantee as to the quality of the winning model.

To solve this problem, Witkowski et al. (2018, 2023) propose the Event Lotteries Forecasting (ELF) mechanism, which is truthful and *accurate*: it selects a good forecaster with high probability when given enough events. Subsequently, Frongillo et al. (2021) show that the Multiplicative Weights algorithm is accurate and more efficient (requires fewer events) than ELF, by relaxing truthfulness to *approximate truthfulness*: forecasters are incentivized to report probabilities close to their beliefs.

All previously proposed mechanisms, however, rely on a strong assumption: that the events in question are independent. By contrast, events in real world forecasting settings like elections, tournaments, and sequential processes are inherently correlated. Swing states tend to swing in the same direction. When a one-seed in a playoff bracket is eliminated early, the overall winner is likely to change. The closing price of a stock today will influence its price tomorrow. To be practical for actual forecasting competitions, mechanisms must be robust to some correlation. Finding a competition mechanism which is provably robust to correlation has therefore been a central open problem in this literature.

In this paper, we prove the first accuracy and approximate truthfulness guarantees for forecasting competitions with correlated events. Unfortunately, as we demonstrate with several examples, not all correlation structures allow a good forecaster to be chosen with high probability (§ 1.1). Moreover, intuitive notions based on covariance or entropy do not seem to adequately quantify the difficulty of forecaster selection. We instead introduce a new way to quantify correlation, called (b, c)-block correlation, and show that it captures several desirable settings (§ 3.1).

Using (b, c)-block correlation, we present a more general accuracy guarantee for Multiplicative Weights which degrades gracefully in the presence of correlation. Specifically, we show that with n forecaters, Multiplicative Weights chooses an ϵ-good forecaster with high probability given $m = O(b^2 \log(n)/\epsilon^2)$ events whose distribution has (b, c)-block correlation for $c = O(\epsilon)$. This bound tightly matches the $O(\log(n)/\epsilon^2)$ bound of (Frongillo et al. 2021) in the independent case, where $b = 1$ and $c = 0$.

There are two keys parts to this proof. In § 4, we show that Multiplicative Weights is still approximately truthful under block correlation, so forecasters will not report values that are much different than their beliefs. Then, in § 5, we use a novel concentration inequality to show our main result, Theorem 2, that with high probability the mechanism M^*_η, Multiplicative Weights with a particular parameter η, efficiently chooses a forecaster with good beliefs. This holds not just in equilibrium of the induced competition game, but whenever forecasters play undominated strategies.

Theorem 2. *For $\eta = \frac{\epsilon}{80b}$, M^*_η chooses an ϵ-optimal forecaster with probability at least $1-\delta$ if there are $m \geq m^* = \frac{400b^2 \ln(8n/\delta)}{\epsilon^2}$ events with $(b, \frac{\epsilon}{20})$-block correlation. In other words, when forecasters play undominated strategies, M^*_η only requires m^* events to choose an ϵ-good forecaster.*

The concentration bound we present for block correlated distributions follows a somewhat complex argument that constructs a pair of connected martingales. It is presented on its own in § 6. We believe this result may be of broader interest; we discuss related literature in § 6.

In summary then, our main contributions are four-fold:

1. Introduction of (b, c)-block correlation.
2. Extension of approximate-truthfulness to the (b, c)-block correlated case.
3. Novel concentration bound for (b, c)-block correlated random variables.
4. Combination of the above to prove Theorem 2.

1.1 How to Capture Correlation?

As Aldous (2003) speculated Tolstoy might say, "All independent variables are alike, all dependent variables are dependent in their own way." Our goal is to define a measure of correlation that quantifies the feasibility of the forecaster selection problem. In this subsection, we walk through a few toy examples which motivate our measure, (b, c)-block correlation.

Single Event. When we have just a single event, there are only two possible outcomes, 0 or 1. Here there are two problems: there is not enough information for the evaluation of forecasters to "concentrate"; and forecasters are incentivized to be highly non-truthful. But general correlation structures can exhibit the same issues. This paper will need to solve both problems in general for the structures we consider. To illustrate the incentive problem, imagine a forecaster who correctly believes the true probability of 1 is 0.5, but also knows there are others predicting 0.1 and 0.9. A competition mechanism will pick one of the others; e.g. if the outcome is 1, then 0.9 appears to be the best forecast. So the optimal response is to "guess" or extremize predictions. Unless the mechanism is designed carefully, this incentive problem does not generally go away as the number of events increases (Witkowski et al. 2023). Moreover, forecasters exploit this issue in real competitions (Kaggle 2017).

Perfectly Correlated Events. Next, let us see why the problem is intractable when correlation is arbitrary. Suppose we have a set of m binary events that are perfectly correlated. Thus, their outcomes will either be all 0 or all 1. With just the single observation, we are reduced to the single event setting where there is not enough information to pick a good forecaster, regardless of m.

Disjoint Correlated Blocks. Instead of all m events being perfectly correlated, suppose they are split into blocks of size b, where each block is independent of all the others. Now, we are essentially in the independent case with m/b events, which is tractable for large enough m. What makes this case tractable while the previous one was not? Instead of just one "underlying" event, we now have m/b of them, and there is enough information to distinguish the forecasters. One might expect that this difference is explained by the amount of randomness, e.g. Shannon entropy of the distributions, since this example has much more than the previous one. However, next we construct distributions with relatively high entropy but only a few "underlying" events.

Random Bias. Suppose all the events are conditionally independent with the same probability p, but p is chosen with equal likelihood from $\{1/4, 3/4\}$. A priori, we will not know p, so the true probability of each event is $1/2$. However, a perfect forecaster who reports $1/2$ for every event will always look worse than one of two competitors who report all $1/4$ or all $3/4$. This looks much like the single-event example, since there is only really one "underlying" event that matters, the choice of p. Yet, the entropy of the outcome distribution is quite high, since every event individually has some randomness. Entropy is therefore not a useful measure for when best-forecaster selection is feasible. We therefore seek a stricter measure that distinguishes the last two examples, which leads us to (b, c)-block correlation.

1.2 Our Measure: (b, c)-block correlation

We capture a more general measure by (b, c)-block correlation (Definition 8), for $b \in \mathbb{N}$ and $c \in [0, 1]$. Binary variables $Y_1, \ldots, Y_m$ are (b, c)-block correlated if for each Y_t, there is a set of at most b variables that "heavily influence" it. Conditioning on all other "non-influencer" events can only change the expectation of Y_t by at most $\pm c$. However, the expectation of Y_t can change arbitrarily when conditioning on any subset of its b "influencer" events.

In the previous random bias example, while there is a lot of randomness, the outcomes of the events are still all strongly coupled. If we know the outcome of a small portion of them, we can probably guess which value of p was chosen, and therefore know the bias of the remaining events with high confidence. Specifically, the random bias example cannot satisfy block correlation with any $c < 1/4$, unless b is very close to m. Knowing just a small fraction of the events can give a strong indication of the choice of p.

In contrast, distributions with favorable block correlation have enough "underlying" events to work with by controlling how many events can strongly influence any one other event. For instance, the Disjoint Correlated Blocks example above satisfies block correlation with $c = 0$ and $b > 0$ for blocks of size b, and can be thought to have m/b underlying events. Our block correlation condition generalizes this example in two respects: it does not require symmetry, i.e. an event can be contained in another's block but not vice versa, and it does not require complete independence outside of the block. We compare to other measures of dependence in the concentration-bounds literature in § 6.

2 Model

We consider n expert forecasters labeled by $[n] = \{1, \ldots, n\}$ predicting the outcomes of m binary events $Y_t \in \{0, 1\}$. Let $Y = (Y_1, \ldots, Y_m)$ be the random variable that is the vector of all the events' outcomes, drawn jointly from the distribution $\mathcal{D}$. Let $\theta = (\mathbb{E}_{\mathcal{D}}[Y_1], \ldots, \mathbb{E}_{\mathcal{D}}[Y_m])$ be the vector of marginal probabilities for each event.

2.1 Beliefs and Accuracy

Each forecaster has their own belief $\mathcal{D}_i$ of what they believe the distribution of Y to be. In particular, we define as $p_i = (\mathbb{E}_{\mathcal{D}_i}[Y_1], \ldots, \mathbb{E}_{\mathcal{D}_i}[Y_m])$ as their belief of θ. While some previous works, such as Witkowski et al. (2023) consider more complex belief models, we assume that the belief distributions are immutable and not influenced by the reports of the other forecasters as in Witkowski et al. (2018) and Frongillo et al. (2021)

To compare forecasters, we use their accuracy, defined as the averaged squared loss between their marginal belief vector p_i and the true marginal probabilities θ. Specifically, the accuracy of forecaster i is given by

$$a_i = 1 - \frac{1}{m} \sum_{t=1}^{m} (p_{it} - \theta_t)^2 \ .$$

As Witkowski et al. (2023) observe, this notion of accuracy is closely related to the quadratic score $S(x, y) = 1 - (x - y)^2$: up to a constant, each forecaster's accuracy is exactly the expected quadratic score between their marginal beliefs and Y.

Lemma 1 (Witkowski et al. (2023))**.** *Let* $C_\theta = \frac{1}{m} \sum_t \theta_t (1 - \theta_t)$*. Then,*

$$a_i = \frac{1}{m} \mathop{\mathbb{E}}_{\mathcal{D}} \left[\sum_{t=1}^{m} S(p_{it}, Y_t) \right] + C_\theta \ .$$

2.2 Selection Mechanisms and Incentives

A forecasting competition mechanism asks each forecaster to report a vector $r_i \in [0, 1]^m$ of what they believe θ to be. If they are truthful, the r_i would be exactly their beliefs p_i. The mechanism then observes a sample $Y = y$ from $\mathcal{D}$, and uses Y along with the reports to choose a forecaster. Let R be the $n \times m$ matrix of combined report vectors $(r_1, \ldots, r_n)$.

Definition 1 (Frongillo et al. (2021))**.** *A **forecasting competition mechanism** M is a family of functions $M_{n,m} : [0, 1]^{n \times m} \times \{0, 1\}^m \to \Delta_n$, for all $n, m \in \mathbb{N}$, where $M_{n,m}(R, y)_i$ is the probability with which the mechanism picks forecaster i on reports R and observed outcomes y. As R determines n and m, we suppress the subscripts. For a particular distribution $\mathcal{D}$, we write $M(R; \mathcal{D}) := \mathbb{E}_{y \sim \mathcal{D}}[M(R, y)]$.*

If forecasters are non-strategic and simply report their beliefs, their accuracy can be approximated by their empirical score (Lemma 1), and choosing a good forecaster is relatively straightforward. In practice, however, forecasters are strategic and may manipulate their reports as a function of their beliefs in order to maximize the probability that they win. To understand which mechanisms are robust to these manipulations, following Frongillo et al. (2021), we will work with undominated and strictly dominated reports.

Definition 2. *A report $\hat{r}_i$ strictly dominates r_i if for all R_{-i}, $M(\hat{r}_i, R_{-i}; \mathcal{D}) > M(r_i, R_{-i}; \mathcal{D})$. If such a $\hat{r}_i$ exists, then r_i is strictly dominated. Otherwise, r_i is undominated.*

If a forecaster reports an r_i that is strictly dominated, then they would only increase their win probability by reporting $\hat{r}_i$ instead. Therefore, strategic forecasters will only give undominated reports. These reports are determined by the mechanism, since each forecaster is trying to maximize the probability the mechanism selects them. If the mechanism always keeps the undominated strategies close to a forecaster's beliefs, we call it approximately-truthful.

Definition 3 (Frongillo et al. (2021))**.** *A mechanism M is γ-approximately truthful if for each $\mathcal{D}_i$, (i) there exists an undominated report r_i, and (ii) for all such r_i, $\|r_i - p_i\|_\infty < \gamma$.*

2.3 Optimality and Efficiency

A desirable mechanism will choose forecasters whose beliefs are close to the true distribution compared to the other forecasters. In particular, we will analyze the probability that mechanisms choose an ϵ-optimal forecaster, one whose accuracy is within ϵ of the best.

Definition 4. *Forecaster j is ϵ-optimal if $a_j + \epsilon \geq \max_{i \in [n]} a_i$.*

Furthermore, we seek mechanisms which choose ϵ-optimal forecasters with high probability under the assumption that forecasters do not submit dominated reports. We call such mechanisms *accurate*.

Definition 5. *A mechanism M is (ϵ, δ)-**accurate** in the setting defined by $(n, m, \{\mathcal{D}_i\}_i, \mathcal{D})$ if for all R consisting of undominated reports, with probability at least $1 - \delta$ over event outcomes $y \sim \mathcal{D}$ and $i \sim M(R, y)$, the winner i is ϵ-optimal.*

Finally, we will study how many events are needed for the accuracy guarantee to hold. A mechanism that always chooses ϵ-optimal forecasters is not useful if it requires far more events to do so than we are able to observe. We define a mechanism's *event complexity* to be the minimum number of events needed for it to choose good forecasters with high probability. In general, we seek mechanisms with a small event complexity.

Definition 6. *The **event complexity** of a mechanism M is the function $m^* : \mathbb{N} \times [0, 1] \times [0, 1] \to \mathbb{N}$ such that, for all n, ϵ, δ, the output $m = m^*(n, \epsilon, \delta)$ is the smallest integer such that, for all $(\{\mathcal{D}_i\}_i, \mathcal{D})$, the mechanism M is (ϵ, δ)-accurate in the setting $(n, m, \{\mathcal{D}_i\}_i, \mathcal{D})$.*

3 Correlation

The accuracy guarantees of previously proposed mechanisms have only been established in settings where the events are independently distributed. We would like to study settings with correlated events. However, as we discuss in § 1.1, we cannot allow for arbitrary correlations. Recall that in the extreme case where all the events perfectly correlated, there are only two possible outcomes, and it becomes impossible to choose a good forecaster with high probability regardless of the number of events.

Therefore, we want to only consider correlated distributions where we can choose a winner with high probability. We need some way to discriminate such settings.

Definition 7. *A function $T : \Delta(\{0,1\}^m) \to \mathbb{R}$ is a* tractable correlation measure *if there exists a function f_T such that for every n and $\epsilon, \delta > 0$, if $T(\mathcal{D}) > f_T(n, \epsilon, \delta)$ then $\Pr_{Y\sim\mathcal{D}}[\arg\max_i S(p_i, Y) \in A_\epsilon(p)] > 1 - \delta$ for all beliefs $p \in [0,1]^{n\times m}$, where $A_\epsilon(p)$ is the set of ϵ-good forecasters, and ties in the argmax are broken randomly.*

Essentially, T is a tractable correlation measure if the mechanism that chooses the player with the maximum quadratic score is (ϵ, δ)-accurate in the non-strategic setting for every $\mathcal{D}$ that meets the the threshold f_T. Although this definition is strongly tied to the choice of score, our choice is well motivated since most mechanisms that are used and studied for forecasting are based upon the quadratic score.

Several standard correlation measures can be immediately ruled out by considering the examples in § 1.1. In particular, the most popular measure of correlation, entropy, is not tractable. Given any threshold f_T, we can increase the number of events in the random bias model such that $\mathcal{D}$ has arbitrarily high entropy. However, as previously described, a forecaster with perfect accuracy may never win with high probability. Other common correlation measures like cross correlation suffer from similar pitfalls.

To proceed, we introduce the *(b, c)-block correlation* property to limit correlation both in the size b of the "block" of events that any one event is strongly correlated with, and in the degree c of correlation persisting outside of that block. Each event may have its own block of size b, and these blocks need not be disjoint.

Definition 8. *A distribution D over Y has (b, c)-block correlation if for each event t there is a subset $B_t \subseteq [m]$ such that $|B_t| \le b$ and*

$$\left| \mathop{\mathbb{E}}_D[Y_t] - \mathop{\mathbb{E}}_D\left[Y_t | Y_{\overline{B_t}} = y_{\overline{B_t}}\right] \right| < c\,, \tag{1}$$

for every $y_{\overline{B_t}} \in \{0,1\}^{m-|B_t|}$.

This definition allows each event t to be arbitrarily correlated with the $\le b$ other events in B_t. While it may also be correlated with the events in $\overline{B_t}$, the amount that correlation is restricted by requiring that for any outcome of the events in $\overline{B_t}$ can only change the conditional mean of Y_t by some small amount c. In general we will need $c < \frac{1}{2}$ so we necessarily have $t \in B_t$.

The concentration bound of Theorem 3 implies that (b, c)-block correlation yields a tractable correlation measure $T_b(\mathcal{D}) = 1/c$ where c is the minimum c such that $\mathcal{D}$ has (b, c)-block correlation.

3.1 Examples

Election Prediction To illustrate (b, c)-block correlation in a typical forecasting setting, we consider the local elections for representatives within a country, say the United States. Similar to the random bias example in § 1.1, we can model each race as a coin flip between electing a Democrat and a Republican, each with some unknown bias $\hat{\theta}_t$, centered around some known mean θ_t. At the national level, common factors affect the overall voter turnout for each party, like their presidential nominee or specific issues in each party's platform. We can think of this as a constant term that is added to the bias of all local election, as people across the country are influenced by these same factors. This is captured by c in our block-correlation condition, since conditioning on the outcomes of many other districts, we can estimate the bias by comparing the θ_t and $\hat{\theta}_t$.

In addition to national factors, local forces within each state may draw people to the polls, such as specific ballot measures or state-level elections. These may cause larger changes in voter turnout, but only within that state. For example, the turnouts within two adjacent districts in New York City are probably strongly coupled, but neither will likely affect how voters in Los Angeles behave. Therefore, if we group the elections into blocks by state, with b being the largest number of local elections in any state, we satisfy (b, c)-block correlation.

Sticky Random Walk (b, c)-block can also capture sequential or Markov processes well. As an example, we consider the Sticky Random Walk, which consists of a sequence of binary events $(Y_1, Y_2, \dots)$ such that $\mathbb{E}[Y_1] = \frac{1}{2}$ and for $t > 1$, $\mathbb{E}[Y_t|Y_{t-1} = 0] = \frac{1}{2} - \alpha$ and $\mathbb{E}[Y_t|Y_{t-1} = 1] = \frac{1}{2} + \alpha$. Essentially, each $Y_t = Y_{t-1}$ with probability $\frac{1}{2} + \alpha$, with the parameter α controlling how "sticky" the walk is.

This process naturally fits the block correlation structure, with the block for event t of size $b = 2k + 1$ being $B_t = \{Y_{t-k}, \dots Y_{t+k}\}$. c can be determined for each choice of α and b. For example, if we fix $b = 3$, for any $c > 8\alpha^4 + 2\alpha^2$, the Y_t will have (b, c)-block correlation.

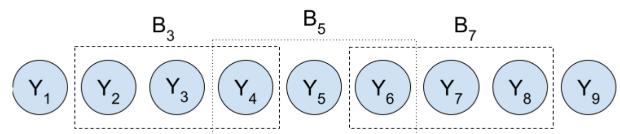

Figure 1: Example blocks for events Y_3, Y_5, and Y_7 of a sticky random walk, where $b = 3$.

3.2 Multiplicative Weights

While Witkowski et al. (2023) show that the truthfulness guarantees of ELF do not hold in the correlated setting, they hypothesize that it may be approximately truthful under mild correlations. However, it is not clear what correlation structures would allow this. Furthermore, Frongillo et al. (2021) show that even in the independent setting, ELF has an event complexity that is $O(n\log(n)/\epsilon^2)$, while Multiplicative Weights' event complexity is only $O(\log(n)/\epsilon^2)$. In correlated settings, we naturally expect ELF's performance to degrade, leaving MW as a more promising mechanism from the perspective of event complexity. Multiplicative Weights is the forecasting competition mechanism M^*_η that chooses forecaster i with probability

$$M^*_\eta(R,y)_i = \frac{\exp\left(\eta \sum_{t=1}^m S(r_{it}, y_t)\right)}{\sum_{j=1}^n \exp\left(\eta \sum_{t=1}^m S(r_{jt}, y_t)\right)} . \quad (2)$$

In the independent setting, Frongillo et al. (2021) show that MW is 4η-approximately truthful and efficiently chooses an ϵ-good forecaster with high probability when $\eta = O(\epsilon)$. Furthermore, they show that the approximate truthfulness guarantees hold for a broad subset of Follow-the-Regularized-Leader (FTRL) mechanisms. We will extend both guarantees for MW to the block correlated setting. We also extend the approximate truthfulness guarantee to the block correlated setting for the same subset of FTRL mechanisms in the extended version of the paper.

4 Approximate Truthfulness of Multiplicative Weights

We begin by showing that MW remains approximately truthful in the presence of correlation. For our analysis, we require that there are constants $b \geq 1, 0 \leq c < \frac{1}{2}$ independent of i such that every $\mathcal{D}_i$ has (b,c)-block correlation. This constraint implies $t \in B_t$ for all t and $\mathcal{D}_i$. We will often refer to B_t as t's block. As all results are special cases of the more general FTRL results proved in the extended version of the paper. We omit their formal proofs here for brevity.

4.1 Utilities

Since each forecaster's objective is to be chosen by the mechanism, their utility is just their expected win probability. Each forecaster seeks to maximize their utility as a function of their report, the only input they have on the mechanism. Fixing the outcomes of the events y and the reports of the other forecasters r_{-i}, forecaster i has utility

$$U_i(r_i) = U_i(r_i; r_{-i}, y) = M^*_\eta(r_i; r_{-i}, y)_i .$$

Taking the expectation over y then gives the expected utility

$$\overline{U}_i(r_i) = \mathop{\mathbb{E}}_{\mathcal{D}_i}\left[U_i(r_i; r_{-i}, Y)\right] .$$

We will also use versions of U_i and $\overline{U}_i$ when we restrict attention to only round t. Specifically, we define

$$\begin{aligned} U_{it}(r_{it}; r_{-(it)}, y) &= U_{it}(r_{it}; r_{i,-t}, r_{-i,-t}, y_{-t}, r_{-i,t}, y_t) \\ &= M^*_\eta(r_{it}; r_{i,-t}, r_{-i,-t}, r_{-i,t}, y)_i \end{aligned}$$

as the utility of forecaster i as a function of their report in round t, with all outcomes and other reports fixed. We use the right hand side to isolate the variables for event t. Now, we can define the expected utility fixing everything but a single event and it's corresponding reports as

$$\overline{U}_{it}(r_{it}) = \mathop{\mathbb{E}}_{\mathcal{D}_{it}}\left[U_{it}(r_{it}; r_{i,-t}, r_{-i,-t}, y_{-t}, r_{-i,t}, Y_t)\right] .$$

Since the events in a block are strongly correlated, it will also be necessary to restrict our attention to the report for a single event t, along with all the outcomes in its block. Fixing any realizations $y_{\overline{B_t}}$ of events not in its block and taking the expectation over the outcomes Y_{B_t} gives us the expected utility

$$\overline{U}_{iBt}(r_{it}) = \mathop{\mathbb{E}}_{\mathcal{D}_i}\left[U_{it}(r_{it}; r_{i,-t}, r_{-i}, y_{\overline{B_t}}, Y_{B_t}) \,|\, Y_{\overline{B_t}} = y_{\overline{B_t}}\right] .$$

4.2 Approximate Truthfulness

In the independent setting, Frongillo et al. (2021) use the strict concavity of U_{it} to find the optimal reports of each forecaster. The idea of their analysis is to fix any outcomes of all events but Y_t. Strict concavity allows them to argue that the optimizer of U_{it} is close to i's unconditional true belief. Since this holds for all realizations of Y_{-t}, it holds in expectation.

With correlated events, their analysis fails because fixing the outcomes of all events but Y_t can cause i's belief, conditional on those outcomes, to change arbitrarily. We modify their approach to only fix the outcomes of events outside of i's block, considering $\overline{U}_{iBt}(r_{it})$. Ignoring c for a moment and pretending Y_t is independent of $Y_{\overline{B_t}}$, we will be able to show that the optimal report is close to i's true belief conditioned on $Y_{\overline{B_t}}$, depending on the size b of the block. Then, we can observe that the true belief conditioned on $Y_{\overline{B_t}}$ is c-close to the unconditional true belief.

We start by showing strict concavity of $\overline{U}_{iBt}$.

Lemma 2. *For $\eta < 1$, for all $i \in [n], t \in [m]$, and all R_{-i}, the functions $\overline{U}_{iBt}(r_{it})$ are strictly concave if $\mathcal{D}_i$ has (b,c)-block correlation.*

Since $\overline{U}_{iBt}(r_{it})$ is strictly concave, it is uniquely maximized at a point. Therefore, fixing everything but the outcomes of the events in B_t, forecaster i has a unique best report r^*_{it} for event t that maximizes their expected utility. Next, we show that r^*_{it} is close to p_{it}, so forecasters will be approximately truthful on event t when the outcomes $Y_{\overline{B_t}}$ are fixed.

Lemma 3. *Let forecaster i's belief $\mathcal{D}_i$ have (b,c)-block correlation. Fix all reports but r_{it} and all outcomes but $Y_{\overline{B_t}}$. Let $r^*_{it} = \arg\max_{r_{it}\in[0,1]} \overline{U}_{iBt}(r_{it})$. Then for $0 < \eta < \frac{1}{3b}$, $|r^*_{it} - p_{it}| \leq 3\eta b + (3\eta b)^2 + c \leq (3b+1)\eta + c$.*

Since the events of $\overline{B_t}$ are less correlated with t, conditioning on their outcomes should not influence the optimal report r^*_{it} much, so we obtain approximate truthfulness when conditioning on all events as well.

Theorem 1. *Let $\gamma = (3b+1)\eta + c$. Then M^*_η is γ-approximately truthful for any $\eta < \frac{1}{3b}$ if the distributions $\mathcal{D}_i$ have (b,c)-block correlation.*

5 Efficiency of Multiplicative Weights

The approximate-truthfulness of M^*_η guarantees that the forecasters reports reflect their beliefs. Next, we will show that this approximate truthfulness implies that the mechanism is accurate and efficient: it will choose a winner with ϵ-good beliefs with high probability, and it only requires $O(b^2 \log(n)/\epsilon^2)$ events to do so.

Both Witkowski et al. (2023) and Frongillo et al. (2021) require that, in addition to the forecasters believing the events were independent, the true distribution of the events adhered to that independence as well. Similarly, in addition to requiring each of the beliefs $\mathcal{D}_i$ to have (b,c)-block correlation, we assume that the true distribution of the outcomes $\mathcal{D}$ does as well.

Recall that $q_i = \sum_t S(r_{it}, y_t)$ is the total quadratic score of forecaster i and that M^*_η takes the q_i as inputs and uses them to choose a winner w according to the distribution given in eq. (2). Frongillo et al. (2021) showed that with high probability, q_w will be close to $\max_i q_i$. Their proof does not depend on $\mathcal{D}$, since it is simply a property of the mechanism, so it applies in the block correlated case as well.

Lemma 4 (Frongillo et al. (2021))**.** *With probability at least $1 - \frac{\delta}{2}$, the winner $w \in [n]$ chosen by M^*_η satisfies $q_w \geq \max_i q_i - \frac{\log(2n/\delta)}{\eta}$.*

This means that Multiplicative Weights chooses a forecaster with good reports, but we want to show that it chooses a forecaster with a high accuracy. Let $q^*_i = \mathbb{E}_{\mathcal{D}}\left[\sum_t S(p_{it}, y_t)\right] = m(a_i - C_\theta)$ be the quadratic scores of each forecaster's beliefs. The last equality follows from Lemma 1. By approximate truthfulness and the Lipshitz properties of the quadratic score, q^*_i must be close to $\mathbb{E}_{\mathcal{D}}[q_i] = \mathbb{E}_{\mathcal{D}}\left[\sum_t S(r_{it}, y_t)\right]$, the expected score of forecaster i's reports. Therefore, if q_i is close to its mean $\mathbb{E}_{\mathcal{D}}[q_i]$ with high probability, then it will also be close to q^*_i.

In the independent setting, Frongillo et al. (2021) show that the q_i concentrate around their means using a straightforward Hoeffding bound. In a correlated setting, this approach will not work, as Hoeffding relies heavily on independence. Instead, we develop a novel concentration bound for block correlated distributions, given by Theorem 3. This bound is complex and interesting in its own right, so we defer its presentation and proof to § 6.

Lemma 5. *If $\mathcal{D}$ has (b, c)-block correlation, then for any i,*

$$\Pr\left[\left|q_i - \mathop{\mathbb{E}}_{\mathcal{D}}[q_i]\right| \geq mc + 2b\sqrt{2m\ln(8n/\delta)}\right] \leq \frac{\delta}{2n}\ .$$

Combining the previous results and choosing η carefully, we can show that Multiplicative Weights efficiently chooses an accurate forecaster even in the presence of correlation.

Theorem 2. *For $\eta = \frac{\epsilon}{80b}$, M^*_η chooses an ϵ-optimal forecaster with probability at least $1 - \delta$ if there are $m \geq m^* = \frac{400b^2\ln(8n/\delta)}{\epsilon^2}$ events with $(b, \frac{\epsilon}{20})$-block correlation. In other words, when forecasters play undominated strategies, M^*_η only requires m^* events to choose an ϵ-good forecaster.*

6 Concentration Bound

The main result of this section is the following concentration bound for sums of (b, c)-block correlated random variables. First, we briefly discuss related literature, focusing on the main pointers. One approach for dependent variables, due to Janson (2004), is to draw a dependency graph among the variables, then derive Hoeffding-like bounds that depend on the fractional chromatic number of the graph. Another general approach is for Markov chains (Lezaud 1998), in which case bounds can be derived depending on spectral properties of the chain (i.e. related to mixing time). A final general approach, as in Kontorovich and Ramanan (2008); Chazottes et al. (2007), involves a "spatial dependency" or "local similarity" assumption on the variables.

However, we do not find these techniques, nor similar dependent-variable bounds such as in (Pelekis and Ramon 2017), sufficient for the forecasting setting we wish to study. For example, (b, c)-block correlation for any $c > 0$ allows for a fully-connected dependency graph (no variables are independent of any subset of the others), so the chromatic-number approach does not apply. We also may not have a Markov-chain structure on the events, for example in a basketball tournament or presidential election. More generally, the above works show sharp concentration of the relevant quantity precisely around its mean, which is not possible with (b, c)-block correlated variables; in general, one can only show concentration to the interval consisting of the mean $\pm c$.

Theorem 3. *Let $Z_1, \ldots, Z_m$ be possibly dependent $[0, 1]$-valued random variables. For each $i \in [m]$, let $B_i \subset [m] \setminus \{i\}$ and $\bar{B}_i := [m] \setminus (B_i \cup \{i\})$. Define $\beta_i := (Z_j : j \in B_i)$ and $\bar{\beta}_i := (Z_j : j \in \bar{B}_i)$. If there is an integer $b \geq 1$ and a constant $c \geq 0$ such that, for all $i \in [m]$, $|B_i| \leq b - 1$ and $\left|\mathbb{E}[Z_i \mid \bar{\beta}_i] - \mathbb{E}[Z_i]\right| \leq c$, then*

$$\Pr\left[\left|\sum_{j=1}^{m} Z_j - \sum_{j=1}^{m} \mathbb{E}[Z_j]\right| \geq mc + 2b\sqrt{2m\ln(4/\delta)}\right] \leq \delta,$$

for all $0 < \delta \leq 1$.

Here, we explain the key idea of the construction that allows us to prove this result. The complete proof is available in the extended version of the paper. For simplicity, consider the case where $B_1 = \{1\}$ and $B_i = \{i-1, i\}$ for $i \geq 2$, i.e. each variable is strongly correlated only with the previous variable.

We would like to treat $S = Z_1 + \cdots + Z_m$ as a martingale sum and use a standard Azuma-Hoeffding approach (e.g (Mitzenmacher and Upfal 2005)). Formally, one bounds $e^{\lambda S} = e^{\lambda Z_1} \cdots e^{\lambda Z_m}$ for some constant λ. The first step would be to fix all realizations of $Z_1, \ldots, Z_{m-1}$ and bound $\mathbb{E}[e^{\lambda Z_m} \mid Z_1, \ldots, Z_{m-1}]$. This is immediate in the standard martingale setting because there, Z_m is independent conditioned on $Z_1, \ldots, Z_{m-1}$. In our case, we could try to "pull in" the influencer variable Z_{m-1} and bound $\mathbb{E}[e^{\lambda Z_m} e^{\lambda Z_{m-1}} \mid Z_1, \ldots, Z_{m-2}]$. Now Z_m is conditionally distributed approximately independently (up to c) and can be controlled; but Z_{m-1} cannot.

Our solution is to add additional variables. Specifically, we create a copy of Z_{m-1} conditioned on $Z_1, \ldots, Z_{m-2}, Z_m$. Call the copy Z'_{m-1}. Call its mean μ_{m-1}, a random variable depending on $Z_1, \ldots, Z_{m-2}, Z_m$. Now consider bounding $S' = S + Z'_{m-1} - \mu_{m-1}$. We rearrange so that the first step is to bound $\mathbb{E}[e^{\lambda Z_m} e^{\lambda(Z_{m-1} - \mu_{m-1})} \mid Z_1, \ldots, Z_{m-2}, Z'_{m-1}]$. Now the inner random variables are bounded and conditionally independent of the outer variables! This holds because μ_{m-1} equals the mean of Z_{m-1} conditioned on *all* of the other variables. Further, we observe that the collection $(Z_1, \ldots, Z_{m-2}, Z'_{m-1})$ is distributed identicially to the collection $(Z_1, \ldots, Z_{m-1})$. So we have successfully controlled Z_m and reduced the problem to controlling $Z_1, \ldots, Z_{m-1}$. Recursing eventually gives us a bound on the final sum.

To summarize, we were able to construct a random variable $X = S + E$, where $X = X_1 + \cdots + X_m$ is a martingale with each variable bounded by b (the size of the largest block). Meanwhile, $E = E_1 + \cdots + E_m$ is also a martingale, and its mean is zero. Concretely in the example above, we had $X_m = Z_m + Z_{m-1} - \mu_{m-1}$ and we had $E_m = Z'_{m-1} - \mu_{m-1}$. With some tweaks to the martingale analysis to account for c, we can show that E concentrates around zero and X concentrates around its expectation, which is equal to the expectation of S. This implies that S concentrates around its expectation as desired.

7 Discussion

We conclude with some conceptual points about our model, and future work.

7.1 Nuances of Ground Truth

Recall that in the single event example in § 1.1, the only event is a single fair coin flip who's outcome is either 0 or 1 with probability $\frac{1}{2}$. There are a few ways to think of what that probability represents. In our model, the probability $\frac{1}{2}$ captures inherent randomness in the world; both outcomes could happen. So if a forecaster reports 1, we consider them to be a poor forecaster, since they are very far from the true probability. Alternatively, there is the deterministic view that there are two universes, one where the coin is 0, and the other where it is 1, and the bias of the coin comes from our uncertainty about which universe we are in. A forecaster who predicts 1 may claim to know that we are in the second universe, and therefore should be considered the best. This difference is essentially whether we consider this information "knowable" or whether the truth is inherently random (cf. aleatoric vs epistemic uncertainty).

In the random bias example (§ 1.1), we had a set of events whose true bias was chosen to be $\frac{1}{4}$ or $\frac{3}{4}$ with equal probability. We consider the best forecaster to be one whose belief is $p_{it} = \frac{1}{2}$ for all t, since this is the marginal probability of each event before the bias is chosen, while we consider a forecaster who thinks the true bias is $\frac{1}{4}$ to be much worse. As in the single event case, however, it is possible that they have some external signal or expertise and is certain (and correct) about the bias. Fundamentally, our high probability guarantees rely on the true distribution $\mathcal{D}$ capturing all information which is knowable, and leaving only randomness which is inherent to the world.

7.2 Measuring Forecaster Accuracy

We measure the accuracy of forecasters by summing the accuracy of their marginal probability of each event. This measure is natural when events are independently distributed, but it is less clear that it is the right measure in the presence of correlation. Under the measure we adopt, if we duplicate an event to two perfectly correlated ones, the relative contribution of that event to a forecaster's accuracy increases. Consequently, in a disjoint block setting (§ 1.1) where blocks may have different sizes, it would perhaps be better to normalize within each block to weigh them evenly. On the other extreme, there are other metrics to compare the entire joint distributions over events, like KL-divergence or total variation distance, yet these would seem to require forecasters to report 2^m probabilities. It would be interesting to explore the space in between, toward accuracy measures that more natively capture correlation but remain tractable.

7.3 Future Work

While we show that FTRL mechanisms are approximately truthful, we only show that Multiplicative Weights is efficient. It remains to be seen if these results apply to FTRL with regularizers other than negative entropy. We would also like to know if there are practical (exactly) truthful mechanisms that are robust to correlation. While one could ask forecasters to report a joint distribution on all events (as a single "meta event") and then use the 1-event version of ELF (which Witkowski et al. (2023) show is still truthful), this mechanism is far from practical in many respects.

Furthermore, while our results show that block correlation is a sufficient condition for efficient forecasting mechanisms, we do not know to what extent it is necessary. It is possible that there are distributions with block correlation parameters that are unfavorable for our analysis, yet still allow for efficient mechanisms. It would be interesting to see what those distributions look like, and if there is a more general property that encompasses those distributions as well as the block correlated ones that work well with our results.

The belief model we adopt could be generalized in several respects. We require that all forecasters believe the true distribution is (b, c)-block correlated for the same constants b and c. In true forecasting competitions, the participants can have a very wide range of beliefs. Even if the true distribution is block correlated, it is possible that there is an extremely misinformed forecasters that believes all the events are always perfectly correlated. It is not clear if the mechanisms we analyzed are robust to such cases. Finally, Witkowski et al. (2023) use a Bayesian model where forecasters may believe that the reports of their competitors are correlated with the truth. For example, when forecasting the weather, we expect meteorologists to make better predictions than most average citizens. We believe that our results will extend to this setting as well.

7.4 Sequential Predictions

Though Frongillo et al. (2021) show the robustness of FTRL in the strategic online setting as well, we focus only on the offline setting. In their online setting, as in Dawid and Tewari (2020) and Choe and Ramdas (2021), forecasters see the outcome of each event before predicting the subsequent one. They can therefore update their belief by conditioning on the events that have already occurred, making their subsequent prediction independent of them, and thus alleviating some of the challenges of correlated mechanisms. The guarantees of Frongillo et al. (2021) in the online setting still hold when accuracy is defined over those conditional distributions.

Acknowledgements

We thank Eduardo Corona, Rupert Freeman, Ziyu Li, David Pennock, Aaditya Ramdas, Ambuj Tewari, and Jens Witkowski, for their helpful ideas, suggestions, and comments. This material is based upon work supported by the National Science Foundation under Grant No. IIS-2045347.

References

Aldous, D. J. 2003. Critical behavior of the Erdos-Renyi random graph: probabilistic methods. https://www.stat.berkeley.edu/~aldous/Networks/Old/lec7.ps. Accessed: 4/30/2023.

Aldous, D. J. 2019. A Prediction Tournament Paradox. *The American Statistician*, 1–6.

Chazottes, J.-R.; Collet, P.; Külske, C.; and Redig, F. 2007. Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields*, 137: 201–225.

Choe, Y. J.; and Ramdas, A. 2021. Comparing sequential forecasters. *arXiv preprint arXiv:2110.00115*.

Dawid, A. P.; and Tewari, A. 2020. On learnability under general stochastic processes. *arXiv preprint arXiv:2005.07605*.

Frongillo, R.; Gomez, R.; Thilagar, A.; and Waggoner, B. 2021. Efficient Competitions and Online Learning with Strategic Forecasters. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 479–496.

Janson, S. 2004. Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms*, 24(3): 234–248.

Kaggle. 2017. March Machine Learning Mania, 1st Place Winner's Interview: Andrew Landgraf. http://blog.kaggle.com/2017/05/19/march-machine-learning-mania-1st-place-winners-interview-andrew-landgraf/. Accessed: 6/29/2019.

Kontorovich, L.; and Ramanan, K. 2008. Concentration inequalities for dependent random variables via the martingale method. *Annals of Probability*, 36: 2126–2158.

Lezaud, P. 1998. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, 849–867.

Lichtendahl, K. C., Jr.; and Winkler, R. L. 2007. Probability Elicitation, Scoring Rules, and Competition Among Forecasters. *Management Science*, 53(11): 1745—1755.

Mitzenmacher, M.; and Upfal, E. 2005. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press.

Pelekis, C.; and Ramon, J. 2017. Hoeffding's inequality for sums of dependent random variables. *Mediterranean Journal of Mathematics*, 14: 1–16.

Witkowski, J.; Freeman, R.; Vaughan, J. W.; Pennock, D. M.; and Krause, A. 2018. Incentive-Compatible Forecasting Competitions. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*.

Witkowski, J.; Freeman, R.; Vaughan, J. W.; Pennock, D. M.; and Krause, A. 2023. Incentive-compatible forecasting competitions. *Management Science*, 69(3): 1354–1374.