Buying Data over Time: Approximately Optimal Strategies for Dynamic Data-Driven Decisions

Nicole Immorlica
Microsoft Research, Cambridge, MA, USA
nicimm@microsoft.com

Ian A. Kash
Department of Computer Science, University of Illinois at Chicago, IL, USA
iankash@uic.edu

Brendan Lucier
Microsoft Research, Cambridge, MA, USA
brlucier@microsoft.com

Abstract

We consider a model where an agent has a repeated decision to make and wishes to maximize their total payoff. Payoffs are influenced by an action taken by the agent, but also an unknown state of the world that evolves over time. Before choosing an action each round, the agent can purchase noisy samples about the state of the world. The agent has a budget to spend on these samples, and has flexibility in deciding how to spread that budget across rounds. We investigate the problem of choosing a sampling algorithm that optimizes total expected payoff. For example: is it better to buy samples steadily over time, or to buy samples in batches? We solve for the optimal policy, and show that it is a natural instantiation of the latter. Under a more general model that includes per-round fixed costs, we prove that a variation on this batching policy is a 2-approximation.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Theory of computation → Markov decision processes

Keywords and phrases Online Algorithms, Value of Data, Markov Processes

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.77

Related Version The full version of this paper is available at `https://arxiv.org/abs/2101.07304`.

Acknowledgements Part of this work was done while Ian Kash was at Microsoft Research. He gratefully acknowledges support from the National Science Foundation via award CCF 1934915.

1 Introduction

The growing demand for machine learning practitioners is a testament to the way data-driven decision making is shaping our economy. Data has proven so important and valuable because so much about the current state of the world is *a priori* unknown. We can better understand the world by investing in data collection, but this investment can be costly; deciding how much data to acquire can be a non-trivial undertaking, especially in the face of budget constraints. Furthermore, the value of data is typically not linear. Machine learning algorithms often see diminishing returns to performance as their training dataset grows [22, 10]. This non-linearity is further complicated by the fact that a data-driven decision approach is typically intended to replace some existing method, so its value is relative to the prior method's performance.

As a motivating example for these issues, consider a politician who wishes to accurately represent the opinion of her constituents. These constituents have a position on a policy, say the allocation of funding to public parks. The politician must choose her own position on the policy or abstain from the discussion. If she states a position, she experiences a disutility that is increasing in the distance of her position from that of her constituents. If she abstains, she incurs a fixed cost for failing to take a stance. To help her make an optimal decision she can hire a polling firm that collects data on the participants' positions.

We focus on the dynamic element of this story. In many decision problems, the state of the world evolves over time. In the example above, the opinions of the constituents might change as time passes, impacting the optimal position of the politician. As a result, data about the state of the world becomes stale. Furthermore, many decisions are not made a single time; instead, decisions are made repeatedly. In our example, the politician can update funding levels each fiscal quarter.

When faced with budget constraints on data collection and the issue of data staleness, decisions need to be made about when to collect data and when to save budget for the future, and whether to make decisions based on stale data or apply a default, non-data-driven policy. Our main contribution is a framework that models the impact of such budget constraints on data collection strategies. In our example, the politician has a budget for data collection. A polling firm charges a fixed cost to initiate a poll (e.g., create the survey) plus a fee per surveyed participant. The politician may not have enough budget to hire the firm to survey every constituent every quarter. Should she then survey fewer constituents every quarter? Or survey a larger number of constituents every other quarter, counting on the fact that opinions do not drift too rapidly?

We initiate the study with arguably the simplest model that exhibits this tension. The state of the world (constituents' opinions) is hidden but drawn from a known prior distribution, then evolves stochastically. Each round, the decision-maker (politician) can collect one or more noisy samples that are correlated with the hidden state at a cost affine in the number of samples (conduct a poll). Then she chooses an action and incurs a loss. Should the decision-maker not exhaust her budget in a given round, she can bank it for future rounds. A sampling algorithm describes an online policy for scheduling the collection of samples given the budget and past observations.

We instantiate this general framework by assuming Gaussian prior, perturbations and sample noise.[1] We capture the decisions that need to be made as the problem of estimating the current state value, using the classic squared loss to capture the cost of making a decision using imprecise information. Alternatively, there is always the option to not make a decision based on the data and instead accept a default constant loss. We assume a budget on the number of samples collected per unit time, and importantly this budget can be banked for future rounds if desired.

1.1 A Simple Example

To illustrate our technical model, suppose the hidden state (constituents' average opinion) is initially drawn from a mean-zero Gaussian of variance 1. In each round, the state is subject to mean-zero Gaussian noise of variance 1 (the constituents update their opinions), which is added to the previous round's state. Also, any samples we choose to take are also subject to mean-zero Gaussian noise of variance 1 (polls are imperfect). Our budget for samples is 1 per period, and one can either guess at the hidden state (incurring a penalty equal to the squared loss) or pass and take a default loss of 3/4. What is the expected average loss of the policy that takes a single sample each round, and then takes the optimal action? As it turns out, the expected loss is precisely $\phi - 1 \approx 0.618$, where ϕ is the golden ratio $\frac{1+\sqrt{5}}{2}$ (see Section 3.5 for the analysis). However, this is not optimal: saving up the allotted budget and taking

[1] A Gaussian prior is justified in our running example if we assume a large population limit of constituents' opinions. That the prior estimate of drift is also Gaussian is likewise motivated as the number of periods grows large. We discuss alternative distributional assumptions on the prior, perturbations and noise in Section 6.

two samples every other round leads to an expected loss of $\frac{0.75+\sqrt{2}-1}{2} \approx 0.582$. The intuition behind the improvement is that taking a single sample every round beats the outside option, but not by much; it is better to beat the outside option significantly on even-numbered rounds (by taking 2 samples), then simply use the outside option on odd-numbered rounds. It turns out that one cannot improve on this by saving up for 3 or more rounds to take even more samples all at once. However, one can do better by alternating between taking no samples for two periods and then two samples each for two periods, which results in a long-run average loss of ≈ 0.576.

1.2 Our Results

As we can see from the example above, the space of policies to consider is quite large. One simple observation is that since samples become stale over time it is never optimal to collect samples and then take the outside option (i.e., default fixed-cost action) in the same round; it would be better to defer data collection to later rounds where decisions will be made based on data. As a result, a natural class of policies to consider is those which alternate between collecting samples and saving budget. Such "on-off" policies can be thought of as engaging in "data drives" while neglecting data collection the rest of the time.

Our main result is that these on-off policies are asymptotically optimal, with respect to all dynamic policies. Moreover, it suffices to collect samples at a constant rate during the sampling part of the policy's period. Our argument is constructive, and we show how to compute an asymptotically optimal policy. This policy divides time into exponentially-growing chunks and collects data in the latter end of each chunk.

The solution above assumes that costs are linear in the number of samples collected. We next consider a more general model with a fixed up-front cost for the first sample collected in each round. This captures the costs associated with setting up the infrastructure to collects samples on a given round, such as hiring a polling firm which uses a two-part tariff. Under such per-round costs, it can be suboptimal to sample in sequential periods (as in an on-off policy), as this requires paying the fixed cost twice. For this generalized cost model, we consider simple and approximately optimal policies. When evaluating performance, we compare against a null "baseline" policy that eschews data collection and simply takes the outside option every period. We define the value of a policy to be its improvement over this baseline, so that the null policy has a value of 0 and every policy has non-negative value. While this is equivalent to simply comparing the expected costs of policies this alternative measure is intended to capture how well a policy leverages the extra value obtainable from data; we feel that this more accurately reflects the relative performance of different policies.

We focus on a class of *lazy policies* that collect samples only at times when the variance of the current estimate is worse than the outside option. This class captures a heuristic based on a threshold rule: the decision-maker chooses to collect data when they do not have enough information to gain over the outside option. We show the optimal lazy policy is a 1/2-approximation to the optimal policy. The result is constructive, and we show how to compute an asymptotically optimal lazy policy. Moreover, this approximation factor is tight for lazy policies.

To derive these results, we begin with the well-known fact that the expected loss under the squared loss cost function is the variance of the posterior. We use an analysis based on Kalman filters [23], which are used to solve localization problems in domains such as astronautics [27], robotics [34], and traffic monitoring [36], to characterize the evolution of variance given a sampling policy. We show how to maximize value using geometric arguments and local manipulations to transform an optimal policy into either an on-off policy or a lazy policy, respectively.

We conclude with two extensions. We described our results for a discrete-time model, but one might instead consider a continuous-time variant in which samples, actions, and state evolution occur continuously. We show how to extend all of our results to such a continuous setting. Second, we describe a non-Gaussian instance of our framework, where the state of the world is binary and switches with some small probability each round. We solve for the optimal policy, and show that (like the Gaussian model) it is characterized by non-uniform, bursty sampling.

1.3 Other Motivating Examples

We motivated our framework with a toy example of a politician polling his or her constituents. But we note that the model is general and applies to other scenarios as well. For example, suppose a phone uses its GPS to collect samples, each of which provides a noisy estimate of location (reasonably approximated by Gaussian noise). The "cost" of collecting samples is energy consumption, and the budget constraint is that the GPS can only reasonably use a limited portion of the phone's battery capacity. The worse the location estimate is, the less useful this information is to apps; sufficiently poor estimates might even have negative value. However, as an alternative, apps always have the outside option of providing location-unaware functionality. Our analysis shows that it is approximately optimal to extrapolate from existing data to estimate the user's location most of the time, and only use the GPS in "bursts" once the noise of the estimate exceeds a certain threshold. Note that in this scenario the app never observes the "ground truth" of the phone's location. Similarly, our model might capture the problem faced by a firm that runs user studies when deciding which features to include in a product, given that such user studies are expensive to run and preferences may shift within the population of customers over time.

1.4 Future Directions

Our results provide insight into the trade-offs involved in designing data collection policies in dynamic settings. We construct policies that navigate the trade-off between cost of data collection and freshness of data, and show how to optimize data collection schedules in a setting with Gaussian noise. But perhaps our biggest contribution is conceptual, in providing a framework in which these questions can be formalized and studied. We view this work as a first step toward a broader study of the dynamic value of data. An important direction for future work is to consider other models of state evolution and/or sampling within our framework, aimed at capturing other applications. For example, if the state evolves in a heavy-tailed manner, as in the non-Gaussian instance explored in Section 6, then we show it is beneficial to take samples regularly in order to detect large, infrequent jumps in state value, and then adaptively take many samples when such a jump is evident. We solve this extension only for a simple two-state Markov chain. Can we quantify the dynamic value of data and find an (approximately) optimal and simple data collection policy in a general Markov chain?

1.5 Related work

While we are not aware of other work addressing the value of data in a dynamic setting, there has been considerable attention paid to the value of data in static settings. Arietta-Ibarra et al. [4] argue that the data produced by internet users is so valuable that they should be compensated for their labor. Similarly, there is growing appreciation for the value of the data produced on crowdsourcing platforms like Amazon Mechanical Turk [6, 20]. Other work has emphasized that not all crowdsourced data is created equal and studied the way tasks and incentives can be designed to improve the quality of information gathered [17, 30]. Similarly, data can have non-linear value if individual pieces are substitutes or complements [8]. Prediction markets can be used to gather information over time, with participants controlling the order in which information is revealed [11].

There is a growing line of work attempting to determine the marginal value of training data for deep learning methods. Examples include training data for classifying medical images [9] and chemical processes [5], as well as for more general problems such as estimating a Gaussian distribution [22]. These studies consider the static problem of learning from samples, and generally find that additional training data exhibits decreasing marginal value. Koh and Liang [25] introduced the use of influence functions to quantify how the performance of a model depends on individual training examples.

While we assume samples are of uniform quality, other work has studied agents who have data of different quality or cost [29, 7, 16]. Another line studies the way that data is sold in current marketplaces [32], as well as proposing new market designs [28]. This includes going beyond markets for raw data to markets which acquire and combine the outputs of machine learning models [33].

Our work is also related to statistical and algorithmic aspects of learning a distribution from samples. A significant body of recent work has considered problems of learning Gaussians using a minimal number of noisy and/or adversarial samples [21, 13, 14, 26, 15]. In comparison, we are likewise interested in learning a hidden Gaussian from which we obtain noisy samples (as a step toward determining an optimal action), but instead of robustness to adversarial noise we are instead concerned about optimizing the split of samples across time periods in a purely stochastic setting.

Our investigation of data staleness is closely related to the issue of concept drift in streaming algorithms; see, e.g., Chapter 3 of [2] Concept drift refers to scenarios where the data being fed to an algorithm is pulled from a model that evolves over time, so that, for example, a solution built using historical data will eventually lose accuracy. Such scenarios arise in problems of histogram maintenance [18], dynamic clustering [3], and others. One problem is to quantify the amount of drift occurring in a given data stream [1]. Given that such drift is present, one approach to handling concept drift is via sliding-window methods, which limit dependence on old data [12]. The choice of window size captures a tension between using a lot of stale data or a smaller amount of fresh data. However, in work on concept drift one typically cannot control the rate at which data is collected.

Another concept related to staleness is the “age of information.” This captures scenarios where a source generates frequent updates and a receiver wishes to keep track of the current state, but due to congestion in the transmission technology (such as a queue or database locks) it is optimal to limit the rate at which updates are sent [24, 31]. Minimizing the age of information can be captured as a limit of our model where a single sample suffices to provide perfect information. Recent work has examined variants of the model where generating updates is costly [19], but the focus in this literature is more on the management of the congestible resource. Closer to our work, several recent papers have eliminated the congestible resource and studied issues such as an energy budget that is stochastic and has limited storage capacity [37] and pricing schemes for when sampling costs are non-uniform [35, 38]. Relative to our work these papers have simpler models of the value of data and focus on features of the sampling policy given the energy technology and pricing scheme, respectively.

2 Model

We first describe our general framework, then describe a specific instantiation of interest in Section 2.1. Time occurs in rounds, indexed by $t = 1, 2, \ldots$. There is a hidden state variable $x_t \in \Omega$ that evolves over time according to a stochastic process. The initial state x_1 is drawn from known distribution F_1. Write m_t for the (possibly randomized) evolution mapping applied at round t, so that $x_{t+1} \leftarrow m_t(x_t)$.

In every round, the decision-maker chooses an action $y_t \in A$, and then suffers a loss $\ell(y_t, x_t)$ that depends on both the action and the hidden state. The evolution functions (m_t) and loss function ℓ are known to the decision-maker, but neither the state x_t nor the loss $\ell(y_t, x_t)$ is directly observed.[2] Rather, on each round before choosing an action, the decision-maker can request one or more independent samples that are correlated with x_t, drawn from a known distribution $\Gamma(x_t)$.

Samples are costly, and the decision-maker has a budget that can be used to obtain samples. The budget is B per round, and can be banked across rounds. A sampling policy results in a number of samples s_t taken in each round t, which can depend on all previous observations. The cost of taking s_t samples in round t is $C(s_t) \geq 0$. We assume that C is non-decreasing and $C(0) = 0$. A sampling policy is *valid* if $\sum_{t=1}^{T} C(s_t) \leq B \cdot T$ for all T. For example, $C(s_t) = s_t$ corresponds to a cost of 1 per sample, and setting $C(s_t) = s_t + z \cdot \mathbb{1}_{s_t > 0}$ adds an additional cost of z for each round in which at least one sample is collected.

To summarize: on each round, the decision-maker chooses a number of samples s_t to observe, then chooses an action y_t. Their loss $\ell(y_t, x_t)$ is then realized, the value of x_t is updated to x_{t+1}, and the process proceeds with the next round. The goal is to minimize the expected long-run average of $\ell(y_t, x_t)$, in the limit as $t \to \infty$, subject to $\sum_{t=0}^{T} C(s_t) \leq B \cdot T$ for all $T \geq 1$.

2.1 Estimation under Gaussian Drift

We will be primarily interested in the following instantiation of our general framework. The hidden state variable is a real number (i.e., $\Omega = \mathbb{R}$) and the decision-maker's goal is to estimate the hidden state in each round. The initial state is $x_1 \sim N(0, \rho)$, a Gaussian with mean 0 and variance $\rho > 0$. Moreover, the evolution process m_t sets $x_{t+1} = x_t + \delta_t$, where each $\delta_t \sim N(0, \rho)$ independently. We recall that the decision-maker knows the evolution process (and hence ρ) but does not directly observe the realizations δ_t.

Each sample in round t is drawn from $N(x_t, \sigma)$ where $\sigma > 0$. Some of our results will also allow fractional sampling, where we think of an $\alpha \in (0, 1)$ fraction of a sample as a sample drawn from $N(x_t, \sigma/\alpha)$.[3] The action space is $A = \mathbb{R} \cup \{\perp\}$. If the decision-maker chooses $y_t \in \mathbb{R}$, her loss is the squared error of her estimate $(y_t - x_t)^2$. If she is too unsure of the state, she may instead take a default action $y_t = \perp$, which corresponds to not making a guess; this results in a constant loss of $c > 0$. Let G_t be a random variable whose law

---

[2] Assuming that the ground truth for $\ell(y_t, x_t)$ is unobserved captures scenarios like our political example, and approximates settings where the decision maker only gets weak feedback, feedback at a delay, or feedback in aggregate over a long period of time. Observing the loss provides additional information about x_{t+1}, and this could be considered a variant of our model where the decision-maker gets some number of samples "for free" each round from observing a noisy version of the loss.

[3] One can view fractional sampling as modeling scenarios where the value of any one single sample is quite small; i.e., has high variance, so that a single "unit" of variance is derived from taking many samples. E.g., sampling a single constituent in our polling example. It also captures settings where it is possible to obtain samples of varying quality with different levels of investment.

is the decision maker's posterior after observing whatever samples are taken in round t as well as all previous samples. The decision maker's subjective expected loss when guessing $y_t \in \mathbb{R}$ is $E[(y_t - G_t)^2]$. This is well known to be minimized by taking $y_t = E[G_t]$, and that furthermore the expected loss is $E[(E[G_t] - G_t)^2] = Var(G_t)$. It is therefore optimal to guess $y_t = E[G_t]$ if and only if $\mathrm{Var}(G_t) < c$, otherwise pass.

We focus on deriving approximately optimal sampling algorithms. To do so, we need to track the variance of G_t as a function of the sampling strategy. As the sample noise and random state permutations are all zero-mean Gaussians, G_t is a zero-mean Gaussian as well, and the evolution of its variance has a simple form.

▶ **Lemma 1.** *Let v_t be the variance of G_t and suppose each $\delta_t \sim N(0, \rho)$ independently, and that each sample is subject to zero-mean Gaussian noise with variance σ. Then, if the decision-maker takes s samples in round $t+1$, the variance of G_{t+1} is*

$$v_{t+1} = \frac{v_t + \rho}{1 + \frac{s}{\sigma}(v_t + \rho)}.$$

The proof, which is deferred to the full version of the paper along with all other proofs, follows from our model being a special case of the model underlying a Kalman filter.

The optimization problem therefore reduces to choosing a number of samples s_t to take in each round t in order to minimize the long-run average of $\min(v_t, c)$, the loss of the optimal action. That is, the goal is to minimize $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \min(v_t, c)$, where we take the superior limit so that the quantity is defined even when the average is not convergent. We choose $C(s_t) = s_t + z \cdot \mathbb{1}_{s_t>0}$, so this optimization is subject to the budget constraint that, at each time $T \geq 1$, $\sum_{t=1}^{T} s_t + z \cdot \mathbb{1}_{s_t>0} \leq BT$. This captures two kinds of information acquisition costs faced by the decision-maker. First she faces a cost per sample, which we have normalized to one. Second, she faces a fixed cost z (which may be 0) on each day she chooses to take samples, expressed in terms of the number of samples that could instead have been taken on some other day had this cost not been paid. This captures the costs associated with setting up the infrastructure to collects samples on a given round, such as getting data collectors to the location where they are needed, hiring a polling firm which uses a two-part tariff, or establishing a satellite connection to begin using a phone's GPS.

A useful baseline performance is the cost of a policy that takes no samples and simply chooses the outside option at all times. We refer to this as the *null policy*. The *value* of a sampling policy s, denoted $\mathrm{Val}(s)$, is defined to be the difference between its cost and the cost of the null policy: $\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \max(c - v_t, 0)$. Note that maximizing value is equivalent to minimizing cost, which we illustrate in Section 3.1. We say that a policy is α-approximate if its value is at least an α fraction of the optimal policy's value.

3 Analyzing Variance Evolution

Before moving on to our main results, we show how to analyze the evolution of the variance resulting from a given sampling policy. We first illustrate our model with a particularly simple class of policies: those where s_t takes on only two possible values. We then analyze arbitrary periodic policies, and show via contraction that they result in convergence to a periodic variance evolution.

3.1 Visualizing the Decision Problem

To visualize the problem, we begin by plotting the result of an example policy where the spending rate is constant for some interval of rounds, then shifts to a different constant spending rate. Figure 1 illustrates one such policy. The spending rates are indicated as

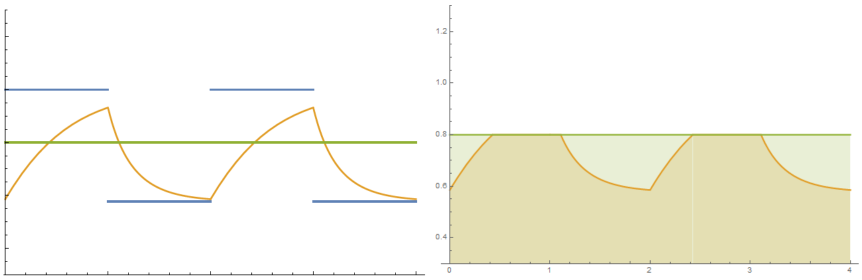

Figure 1 The variance for a piecewise-constant sampling policy, and its loss and benefit.

alternating line segments, while the variance is an oscillating curve, always converging toward the current spending rate. Note that this particular policy is periodic, in the sense that the final variance is the same as the initial variance. The horizontal line gives one possible value for the cost of the outside option. Given this, the optimal policy is to guess whenever the orange curve is below the green line and take the outside option whenever it is above it. Thus, the loss associated with this spending policy is given by the orange shaded area in Figure 1. Minimizing this loss is equivalent to maximizing the green shaded area, which corresponds to the value of the spending policy. The null policy, which takes no samples and has variance greater than c always (possibly after an initial period if $v_0 < c$), has value 0.

3.2 Periodic Policies

We next consider policies that are periodic. A *periodic policy* with period R has the property that $s_t = s_{t+R}$ for all $t \geq 1$. Such policies are natural and have useful structure. In a periodic policy, the variance (v_t) converges uniformly to being periodic in the limit as $t \to \infty$. This follows because the impact of sampling on variance is a contraction map.

▶ **Definition 2.** *Given a normed space X with norm $||\cdot||$, a mapping $\Psi\colon X \to X$ is a* contraction map *if there exists a $k < 1$ such that, for all $x, y \in X$, $||\Psi(x) - \Psi(y)|| \leq k||x - y||$.*

▶ **Lemma 3.** *Fix a sampling policy s, and a time $R \geq 1$, and suppose that s takes a strictly positive number of samples in each round $t \leq R$. Let Ψ be the mapping defined as follows: supposing that $v_0 = x$ and v is the variance function resulting from sampling policy s, set $\Psi(x) := v_R$. Then Ψ is a contraction map over the non-negative reals, under the absolute value norm.*

The proof appears in the full version of the paper. It is well known that a contraction map has a unique fixed point, and repeated application will converge to that fixed point. Since we can view the impact of the periodic sampling policy as repeated application of mapping Ψ to the initial variance in order to obtain $v_0, v_R, v_{2R}, \ldots$, we conclude that the variance will converge uniformly to a periodic function for which $v_t = v_{t+R}$. Thus, for the purpose of evaluating long-run average cost, it will be convenient (and equivalent) to replace the initial condition on v_0 with a periodic boundary condition $v_0 = v_R$, and then choose s to minimize the average cost over a single period, $\frac{1}{R}\int_0^R \min\{v_t, c\}dt$, subject to the budget constraint that, at any round $T \leq R$, we have $\sum_{t=1}^{T} s_t \leq BT$.

3.3 Lazy Policies

Write $\tilde{v} = v_{t-1} + \rho$ for the variance that would be obtained in round t if $s_t = 0$. We say that a policy is *lazy* if $s_t = 0$ whenever $\tilde{v}_t < c$. That is, samples are collected only at times where the variance would otherwise be at or above the outside option value c. Intuitively, we can think of such a policy as collecting a batch of samples in one round, then "free-riding" off of the resulting information in subsequent rounds. The free-riding occurs until the posterior variance grows large enough that it becomes better to select the outside option, at which point the policy may collect another batch of samples.

If a policy is lazy, then its variance function v increases by ρ whenever $\tilde{v}_t < c$, with downward steps only at times corresponding to when samples are taken. Furthermore, the value of such a policy decomposes among these sampling instances: for any t where $s_t > 0$, resulting in a variance of $v_t < c$, if we write $h = \lfloor c - v_t \rfloor$ then we can attribute a value of $\frac{1}{2}h(h+1) + (h+1)(c - v_t - h)$. Geometrically, this is the area of the "discrete-step triangle" formed between the increasing sequence of variances v_t and the constant line at c, over the time steps $t, \ldots, t+h+1$.

3.4 On-Off Policies

An On-Off policy is a periodic policy parameterized by a time interval T and a sampling rate S. Roughly speaking, the policy alternates between intervals where it samples at a rate of S each round, and intervals where it does not sample. The two interval lengths sum to T, and the length of the sampling interval is set as large as possible subject to budget feasibility. More formally, the policy sets $s_t = 0$ for all $t \le (1-\alpha) \cdot T$, where $\alpha = \min\{B/S, 1\} \in [0,1]$ and $s_t = S$ for all t such that $(1-\alpha)T < t \le T$. This policy is then repeated, on a cycle of length T. The fraction α is chosen to be as large as possible, subject to the budget constraint.

3.5 Simple Example Revisited

We can now justify the simple example we presented in the introduction, where $\rho = \sigma = 1$, $B = 1$, and $c = 0.75$. The policy that takes a single sample each round is periodic with period 1, and hence will converge to a variance that is likewise equal each round. This fixed point variance, v^*, satisfies $v^* = \frac{v^*+1}{1+(v^*+1)}$ by Lemma 1. Solving for v^* yields $v^* = \frac{\sqrt{5}-1}{2} < 0.75$, which is the average cost per round.

If instead the policy takes k samples every k rounds, this results in a variance that is periodic of period k. After the round in which samples are taken, the fixed-point variance satisfies $v^* = \frac{v^*+k}{1+k(v^*+k)}$, again by Lemma 1. Solving for v^*, and noting that $v^*+1 \ge 1 > c$, yields that the cost incurred by this policy is minimized when $k = 2$.

To solve for the policy that alternates between taking no samples for two round, followed by taking two samples on each of two rounds, suppose the long-run, periodic variances are v_1, v_2, v_3, v_4, where samples are taken on rounds 3 and 4. Then we have $v_2 = v_1 + 1$, $v_3 = \frac{v_2+1}{1+2(v_2+1)}$, $v_4 = \frac{v_3+1}{1+2(v_3+1)}$, and $v_1 = v_4 + 1$. Combining this sequence of equations yields $4v_1^2 + 4v_1 - 13 = 0$, which we can solve to find $v_1 = \frac{-1+\sqrt{14}}{2} \approx 1.3708$. Plugging this into the equations for v_2, v_3, v_4 and taking the average of $\min\{v_i, 0.75\}$ over $i \in \{1,2,3,4\}$ yields the reported average cost of ≈ 0.576.

4 Solving for the Optimal Policy

In this section we show that when the cost of sampling is linear in the total number of samples taken (i.e., $z = 0$)[4], and when fractional sampling is allowed, then the supremum value over all on-off policies is an upper bound on the value of any policy. This supremum

[4] Recall that z is the fixed per-round cost of taking a positive number of samples. Even when $z = 0$, there is still a positive per-sample cost.

is achieved in the limit as the time interval T grows large. So, while no individual policy achieves the supremum, one can get arbitrarily close with an on-off policy of sufficiently long period. Proofs appear in the full version of the paper.

We begin with some definitions. For a given period length $T > 0$, write s^T for the on-off policy of period T with optimal long-run average value. Recall $\mathrm{Val}(s^T)$ is the value of policy s_T. We first argue that larger time horizons lead to better on-off policies.

▶ **Lemma 4.** *With fractional samples, for all $T > T'$, we have $Val(s^T) > Val(s^{T'})$.*

Write $V^* = \sup_{T\to\infty} \mathrm{Val}(s^T)$. Lemma 4 implies that $V^* = \lim_{T\to\infty} \mathrm{Val}(s^T)$ as well. We show that no policy satisfying the budget constraint can achieve value greater than V^*.

▶ **Theorem 5.** *With fractional samples, the value of any valid policy s is at most V^*.*

The proof of Theorem 5 proceeds in two steps. First, for any given time horizon T, it is suboptimal to move from having variance below the outside option to above the outside option; one should always save up budget over the initial rounds, then keep the variance below c from that point onward. This follows because the marginal sample cost of reducing variance diminishes as variance grows, so it is more sample-efficient to recover from very high variance once than to recover from moderately high variance multiple times.

Second, one must show that it is asymptotically optimal to keep the variance not just below c, but uniform. This is done by a potential argument, illustrating that a sequence of moves aimed at "smoothing out" the sampling rate can only increase value and must terminate at a uniform policy. The difficulty is that a sample affects not only the value in the round it is taken, but in all subsequent rounds. We make use of an amortization argument that appropriately credits value to samples, and use this to construct the sequence of adjustments that increase overall value while bringing the sampling sequence closer to uniform in an appropriate metric.

We also note that it is straightforward to compute the optimal on-off policy for a given time horizon T, by choosing the sampling rate that maximizes [value per round] × [fraction of time the policy is "on"]. One can implement a policy whose value asymptotically approaches V^* by repeated doubling of the time horizon. Alternatively, since $\lim_{T\to\infty} \mathrm{Val}(s^T) = V^*$, s^T will be an approximately optimal policy for sufficiently large T.

5 Approximate Optimality of Lazy Policies

In the previous section we solved for the optimal policy when $z = 0$, meaning that there is no fixed per-round cost when sampling. We now show that for general z, lazy policies are approximately optimal, obtaining at least 1/2 of the value of the optimal policy. Proofs appears in the full version of the paper.

We begin with a lemma that states that, for any valid sampling policy and any sequence of timesteps, it is possible to match the variance at those timesteps with a policy that only samples at precisely those timesteps, and the resulting policy will be valid.

▶ **Lemma 6.** *Fix any valid sampling policy s (not necessarily lazy) with resulting variances (v_t), and any sequence of timesteps $t_1 < t_2 < \ldots < t_\ell < \ldots$. Then there is a valid policy s' such that $\{t \mid s'_t > 0\} \subseteq \{t_1, \ldots, t_\ell, \ldots\}$, resulting in a variances $(\check{v}_t)$ with $\check{v}_{t_i} \le v_{t_i}$ for all i.*

The intuition is that if we take all the samples we would have spent between timesteps t_ℓ and $t_{\ell+1}$ and instead spend them all at $t_{\ell+1}$ the result will be a (weakly) lower variance at $t_{\ell+1}$. We next show that any policy can be converted into a lazy policy at a loss of at most half of its value.

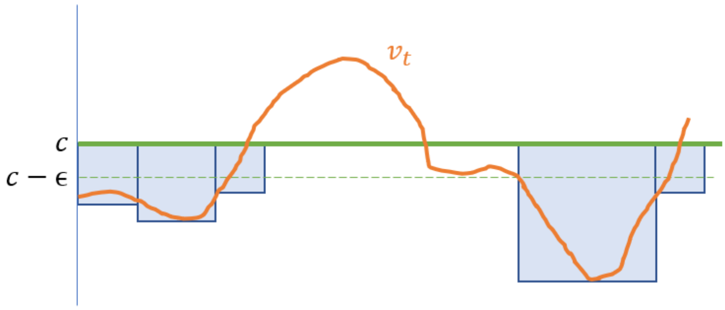

Figure 2 Visualizing the construction in the proof of Theorem 7. Variance (vertical) is plotted against time (horizontal). We approximate the value of an optimal policy's variance (orange) given c (green). The squares (drawn in blue) cover the gap between the curves, except possibly when $|v_t - c| < \epsilon$ (for technical reasons). The lazy policy samples on rounds corresponding to the left edge of each square, bringing the variance to each square's bottom-left corner.

▶ **Theorem 7.** *The optimal lazy policy is* $1/2$*-approximate.*

See Figure 2 for an illustration of the intuition behind the result. Consider an arbitrary policy s, with resulting variance sequence (v_t). Imagine covering the area between (v_t) and c with squares, drawn left to right with their upper faces lying on the outside option line, each chosen just large enough so that v_t never falls below the area covered by the squares. The area of the squares is an upper bound on $\mathrm{Val}(s)$. Consider a lazy policy that drops a single atom on the left endpoint of each square, bringing the variance to the square's lower-left corner. The value of this policy covers at least half of each square. Moreover, Lemma 6 implies this policy is (approximately) valid, as it matches variances from the original policy, possibly shifted early by a constant number of rounds. This shifting can introduce non-validity; we fix this by delaying the policy's start by a constant number of rounds without affecting the asymptotic behavior.

The factor $1/2$ in Theorem 7 is tight. To see this, fix the value of c and allow the budget B to grow arbitrarily large. Then the optimal value tends to c as the budget grows, since the achievable variance on all rounds tends to 0. However, the lazy policy cannot achieve value greater than $c/2$, as this is what would be obtained if the variance reached 0 on the rounds on which samples are taken.

Finally, while this result is non-constructive, one can compute a policy whose value approaches an upper bound on the optimal lazy policy, in a similar manner to the optimal on-off policy. One can show the best lazy policy over any finite horizon has an "off" period (with no sampling) followed by an "on" period (where $v_t \leq c$). One can then solve for the optimal number of samples to take whenever $\tilde{v}_t > c$ by optimizing either value per unit of (fixed plus per-sample) sampling cost, or by fully exhausting the budget, whichever is better. Details appear in the full version of the paper.

6 Extensions and Future Directions

We describe two extensions of our model in the appendix. First, we consider a continuous-time variant where samples can be taken continosly subject to a flow cost, in addition to being requested as discrete atoms. The decision-maker selects actions continuously, and aims to minimize loss over time. All of our results carry forward to this continuous extension.

Second, returning to discrete time, we consider a non-Gaussian instance of our framework.

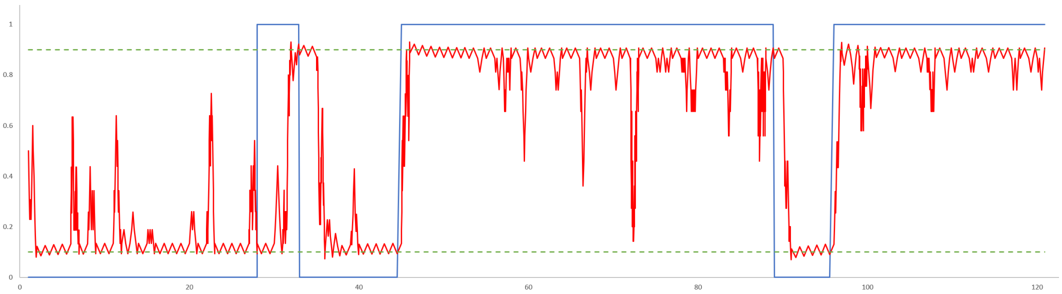

Figure 3 Simulating the optimal policy for the non-Gaussian extension. The round number is on the horizontal axis. The hidden state of the world is binary and evolves stochastically (blue). The optimal policy tracks a posterior distribution over the hidden state (red), and takes samples in order to maintain a tuned level of certainty (dashed green). Note that most rounds have only a small number of samples, with occasional spikes triggered adaptively in response to uncertainty.

In this model, there is a binary hidden state of the world, which flips each round independently with some small probability $\epsilon > 0$. The decision-maker's action in each round is to guess the hidden state of this simple two-state Markov process, and the objective is to maximize the fraction of time that this guess is made correctly. Each sample is a binary signal correlated with the hidden state, matching the state of the world with probability $\frac{1}{2}+\delta$ where $\delta > 0$. The decision-maker can adaptively request samples in each round, subject to the accumulating budget constraint, before making a guess.

In this extension, as in our Gaussian model, the optimal policy collects samples non-uniformly. In fact, the optimal policy has a simple form: it sets a threshold $\theta > 0$ and takes samples until the entropy of the posterior distribution falls below θ. Smaller θ leads to higher accuracy, but also requires more samples on average, so the best policy will set θ as low as possible subject to the budget constraint. Notably, the result of this policy is that sampling tends to occur at a slow but steady rate, keeping the entropy around θ, except for occasional spikes of samples in response to a perceived change in the hidden state. See Figure 3 for a visualization of a numerical simulation with a budget of 6 samples (on average) per round.

More generally, whenever the state evolves in a heavy-tailed manner, it is tempting to take samples regularly in order to detect large, infrequent jumps in state value, and then adaptively take many samples when such a jump is evident. This simple model is one scenario where such behavior is optimal. More generally, can we quantify the dynamic value of data and find an (approximately) optimal data collection policy for more complex Markov chains, or other practical applications?

References

1 Charu C. Aggarwal. A framework for diagnosing changes in evolving data streams. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, SIGMOD '03, pages 575–586, New York, NY, USA, 2003. ACM. `doi:10.1145/872757.872826`.

2 Charu C. Aggarwal. *Data Streams: Models and Algorithms (Advances in Database Systems)*. Springer-Verlag, Berlin, Heidelberg, 2006.

3 Charu C. Aggarwal, Jiawei Han, Jianyong Wang, and Philip S. Yu. A framework for clustering evolving data streams. In *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29*, VLDB '03, pages 81–92. VLDB Endowment, 2003. URL: `http://dl.acm.org/citation.cfm?id=1315451.1315460`.

4 Imanol Arrieta-Ibarra, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E Glen Weyl. Should we treat data as labor? moving beyond" free". In *AEA Papers and Proceedings*, volume 108, pages 38–42, 2018.

5 Claudia Beleites, Ute Neugebauer, Thomas Bocklitz, Christoph Krafft, and Jürgen Popp. Sample size planning for classification models. *Analytica chimica acta*, 760:25–33, 2013.

6 Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

7 Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, and Juba Ziani. Optimal data acquisition for statistical estimation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 27–44. ACM, 2018.

8 Yiling Chen and Bo Waggoner. Informational substitutes. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 239–247. IEEE, 2016.

9 Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *CoRR*, abs/1511.06348, 2015. `arXiv:1511.06348`.

10 Corinna Cortes, Lawrence D Jackel, Sara A Solla, Vladimir Vapnik, and John S Denker. Learning curves: Asymptotic values and rate of convergence. In *Advances in Neural Information Processing Systems*, pages 327–334, 1994.

11 Bo Cowgill, Justin Wolfers, and Eric Zitzewitz. Using prediction markets to track information flows: Evidence from google. In *AMMA*, page 3, 2009.

12 Mayur Datar, Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Maintaining stream statistics over sliding windows: (extended abstract). In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '02, pages 635–644, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics. URL: `http://dl.acm.org/citation.cfm?id=545381.545466`.

13 I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, A. Moitra, and A. Stewart. Robust estimators in high dimensions without the computational intractability. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 655–664, 2016.

14 I. Diakonikolas, D. M. Kane, and A. Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84, 2017.

15 Ilias Diakonikolas, Gautam Kamath, Daniel M. Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '18, pages 2683–2702, Philadelphia, PA, USA, 2018. Society for Industrial and Applied Mathematics. URL: `http://dl.acm.org/citation.cfm?id=3174304.3175475`.

16 Fang Fang, Maxwell Stinchcombe, and Andrew Whinston. " putting your money where your mouth is"-a betting platform for better prediction. *Review of Network Economics*, 6(2), 2007.

17 Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. Instructing people for training gestural interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1737–1746. ACM, 2012.

18 Anna C. Gilbert, Sudipto Guha, Piotr Indyk, Yannis Kotidis, S. Muthukrishnan, and Martin J. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 389–398, New York, NY, USA, 2002. ACM. `doi:10.1145/509907.509966`.

19 Shugang Hao and Lingjie Duan. Regulating competition in age of information under network externalities. *IEEE Journal on Selected Areas in Communications*, 38(4):697–710, 2020.

20 Panagiotis G Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, 2010.

21 Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Efficiently learning mixtures of two gaussians. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 553–562, New York, NY, USA, 2010. Association for Computing Machinery. `doi:10.1145/1806689.1806765`.

22 H. M. Kalayeh and D. A. Landgrebe. Predicting the required number of training samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(6):664–667, November 1983. `doi:10.1109/TPAMI.1983.4767459`.

23 Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

24 Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *2012 Proceedings IEEE INFOCOM*, pages 2731–2735. IEEE, 2012.

25 Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning*, pages 1885–1894, 2017.

26 K. A. Lai, A. B. Rao, and S. Vempala. Agnostic estimation of mean and covariance. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 665–674, Los Alamitos, CA, USA, October 2016. IEEE Computer Society. `doi:10.1109/FOCS.2016.76`.

27 Ern J Lefferts, F Landis Markley, and Malcolm D Shuster. Kalman filtering for spacecraft attitude estimation. *Journal of Guidance, Control, and Dynamics*, 5(5):417–429, 1982.

28 Chao Li and Gerome Miklau. Pricing aggregate queries in a data marketplace. In *WebDB*, pages 19–24, 2012.

29 Annie Liang, Xiaosheng Mu, and Vasilis Syrgkanis. Dynamic information acquisition from multiple sources. *arXiv preprint*, 2017. `arXiv:1703.06367`.

30 Nihar Bhadresh Shah and Denny Zhou. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. In *Advances in neural information processing systems*, pages 1–9, 2015.

31 Xiaohui Song and Jane W-S Liu. Performance of multiversion concurrency control algorithms in maintaining temporal consistency. In *Proceedings Fourteenth Annual International Computer Software and Applications Conference*, pages 132–133. IEEE Computer Society, 1990.

32 Florian Stahl, Fabian Schomm, and Gottfried Vossen. The data marketplace survey revisited. Technical report, Working Papers, ERCIS-European Research Center for Information Systems, 2014.

33 Amos Storkey. Machine learning markets. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 716–724, 2011.

34 Sebastian Thrun. Probabilistic algorithms in robotics. *Ai Magazine*, 21(4):93, 2000.

35 Xuehe Wang and Lingjie Duan. Dynamic pricing for controlling age of information. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 962–966. IEEE, 2019.

36 Daniel B Work, Olli-Pekka Tossavainen, Sébastien Blandin, Alexandre M Bayen, Tochukwu Iwuchukwu, and Kenneth Tracton. An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 5062–5068. IEEE, 2008.

37 Xianwen Wu, Jing Yang, and Jingxian Wu. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Transactions on Green Communications and Networking*, 2(1):193–204, 2017.

38 Meng Zhang, Ahmed Arafa, Jianwei Huang, and H Vincent Poor. How to price fresh data. *arXiv preprint*, 2019. `arXiv:1904.06899`.