Opposite Online Learning via Sequentially Integrated Stochastic Gradient Descent Estimators

Wenhai Cui[1*], Xiaoting Ji[1*], Linglong Kong[3], Xiaodong Yan[124†]

[1] Zhongtai Securities Institute for Financial Studies, Shandong University
[2] Department of Mathematical and Statistical Sciences, University of Alberta
[3] Shandong Province Key Laboratory of Financial Risk
[4]Shandong National Center for Applied Mathematics
{cuiwenhai, jixiaoting}@mail.sdu.edu.cn, lkong@ualberta.ca, yanxiaodong@sdu.edu.cn

Abstract

The stochastic gradient descent algorithm, often depicted as SGD, has been widely employed in various fields of artificial intelligence and is a prototype of online learning algorithms. In the article, we propose a novel and general framework of one-sided testing for streaming data based on SGD. The proposed method constructs an online-updated test statistic sequentially by integrating the selected batch-specific estimators or its opposite, which is referred to as opposite online learning. Notably, the batch-specific online estimators are chosen strategically according to the proposed sequential tactics designed by the two-armed bandit process. Theoretical results prove the strategy's advantage, ensuring that the test statistic distribution is optimal under the null hypothesis. We also supply the theoretical evidence of power enhancement compared with classical test statistics. In application, the proposed method is appealing for statistical inference of two-sided testing and it is scalable and adaptable for any model. Finally, the superior finite-sample performance is evaluated by simulation studies.

Introduction

In recent years, advanced data collection technologies can record streaming data set in clinical, financial, and sociological studies, which promotes the development of some emerging online learning procedure to deal with sequential batches. For the reason that the computer does not have enough memory to store the entire data set, the conventional offline-based methods have been less attractive and no longer applicable.

Online learning procedure, intensively argued in some advanced machine learning methods, has been popularly used to update the population parameter of interest. Among online learning procedures, the stochastic gradient descent (Robbins and Monro 1951) is a scalable algorithm for parameter estimation and has recently drawn a great deal of attention. Unlike other classical methods involving the evaluation of an entire dataset for an objective function, the SGD method uses one data point at a time to compute the gradient of the objective function and recursively updates the parameter estimates. This is also tremendously appealing and particularly useful in online update settings, such as streaming data, where storing the entire dataset in memory may not even be feasible. The asymptotic properties of SGD estimators, such as consistency and asymptotic normality, have been well established; see, for example, Ruppert (1988), Polyak and Juditsky (1992). We can refer to various online methods, including stochastic gradient descent (SGD) and its variants (Chen et al. 2020; Zhu, Chen, and Wu 2021; Liu, Yuan, and Shang 2022). In detail, Langford, Li, and Zhang (2009) proposed a variant of the truncated SGD in online settings. Except for SGD updates, other online learning methods is referred by aggregated estimating equation (Lin and Xi 2011), cumulatively updating estimating equation (Schifano et al. 2016a), renewable estimator (Luo and Song 2019), diffusion approximation approach (Fan et al. 2018). However, the above online algorithms are enlightened by statistical estimation, and there are rare methods that argue hypothesis test.

Therefore, this work was motivated by the issue of statistical inference for streaming data, and we have explored that online updates are closely related to the two-armed bandit (TAB) process which is the prototype of a slot TAB machine (Bellman 1956) with two arms L(left) and R(right), which behaves like obtaining the batch-specific online estimators sequentially. Specifically, at each step with a batch of data set, we play arm L(left) to derive a batch-specific online estimator or obtain its opposite form by using arm R(right), and this procedure is referred to as the opposite online learning. At last, we integrate all batch-specific estimators based on the formation of Chen, Feng, and Zhang (2022) because they have established the central limit theorem (CLT) for the TAB model, which is the primary theoretical result for our considered statistical inference. Although Shi, Song, Lu, and Li (2021) developed an inference procedure for high-dimensional generalized linear models based on recursive online-score estimation, Deshpande, Javanmard, and Mehrabi (2021) considered the high-dimensional linear regression and Han, Luo, Lin, and Huang (2021) proposed an online statistical inference procedure in high-dimensional linear models with streaming data, their established inference procedures are based on conventional theoretical re-

*These authors contributed equally.

†Xiaodong Yan is the corresponding author.

sults of classical normal distribution based central limit theorem. Chen and Epstein (2022) have proved that their proposed central limit theorem for the TAB model is superior to classical CLT because they have used the prior information to construct the test statistic.

In this paper, we propose a novel online learning procedure for both one-sided and two-sided tests of streaming data by using bootstrap-based stochastic gradient descent algorithm (Fan et al. 2018) for estimation and employing the TAB process-based central limit theorem (Chen and Epstein 2022) to integrate all batch-specific estimators.

Our contributions and the merits of the proposed method of opposite online learning are multiple aspects, including

- It is the first attempt using the TAB model, a simple reinforcement learning process, in the standard statistical testing issue.
- We have concluded the asymptotic distribution of the proposed test statistic by the TAB process-based central limit theorem under the null and alternative hypotheses, which lays an essential theoretical tool for statistical inference.
- The proposed testing procedure is scalable for any statistical inference, including one-sided and two-sided tests.
- The proposed opposite online learning method is a generalized method, which can be applied to a broader range of models under streaming data because of the use of stochastic gradient descent.

The rest of this paper is organized as follows. In Section 2, we proposed a bootstrap procedure based on SGD to estimate the parameter of streaming data. In Section 3, we established the corresponding theoretical results which showed more significant power than the traditional test. In Section 4, we proposed a method to test the significance of coefficients based on the above theoretical properties. The finite-sample properties of the proposed method are evaluated through the simulation studies in Section 5. Discussion is provided in Section 6, and the technical proofs are given in the Appendix.

Methodology

Problem

Suppose that T batches of data sets $\{S_1, S_2, \ldots, S_T\}$ are observed in sequence, where the set $S_t = \{(Z_{t,1}, Z_{t,2}, \ldots, Z_{t,N_t})\}$ consists of independent identically distributed samples with the sample size denoted as N_t. Thus, after $T+1$ days, the $T+1$th batch of data set S_t will be incorporated into the set $\mathcal{D}_{T+1} = \{S_1, S_2, \ldots, S_T, S_{T+1}\}$.

Once a loss function is established according to our purposes, prediction or classification, we can search for optimal estimates of parameters via some classical optimization algorithms. Hence, we focus on testing the following pair of hypotheses:

$$\mathbf{H}_0 : \theta_0 \geq d_0; \quad \mathbf{H}_1 : \theta_0 < d_0,$$

where d_0 is a pre-specified constant. The optimal model parameter $\theta_0 \in \mathcal{R}^p$ is denoted as

$$\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmin}}\{L(\theta) \triangleq \mathbb{E}[\ell(\theta; Z)]\}.$$

When θ_0 is a multi-dimensional vector, it is natural to concentrate on every element of θ_0.

Stochastic Gradient Descent for Parameter Estimation

After the loss function $\ell(\theta; Z)$ is set up, it is reasonable to utilize the gradient descent method to solve the model parameters. However, the estimates will fall in local optimal if the loss function is not a convex one. Additionally, the large scale of the data makes it unsurmountable to calculate the gradients of all training data within a tremendously short time. Consequently, we employ SGD to jump out of the local optimal solution and to accelerate the computational efficiency.

Exploiting the random approximation method (Robbins and Monro 1951), the stochastic gradient descent algorithm just uses one observation of the whole data S_t to update parameters per iteration, i.e.,

$$\widehat{\theta}_{n,t} = \widehat{\theta}_{n-1,t} - \gamma_n \nabla\ell\left(\widehat{\theta}_{n-1}; Z_{t,n}\right), \tag{0.1}$$

where the learning rate γ_n is equal to $\gamma_1 n^{-\alpha}$ with $\gamma_1 > 0$ and $\alpha \in (0.5, 1)$. As proposed by Polyak and Juditsky (1992), the averaging estimate is given as,

$$\bar{\theta}_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{i,t}. \tag{0.2}$$

Bootstrap Based Stochastic Gradient Descent

It is well known that stochastic gradient descent is still easy to fall into saddle points and local minimum points. To obtain more robust results, an online bootstrap resampling approach is used to update recursively the SGD estimators at time point t, which means that the the properties of the SGD estimator distribution are considered.

Specifically, let $\mathcal{P} = \{P_i, i = 1, \ldots, B\}$ be a set of i.i.d. non-negative random variables with mean and variance being equal to one. Likewise, once a observed sample $(Z_{t,n})$ arrives, we recursively update the randomly perturbed SGD estimates by

$$\widehat{\theta}^*_{n,t} = \widehat{\theta}^*_{n-1} - \gamma_n P_n \nabla\ell\left(\widehat{\theta}^*_{n-1}; Z_{t,n}\right)$$
$$\bar{\theta}^*_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}^*_i,$$

where $\bar{\theta}^*_t$ denotes the estimate by randomly perturbed SGD algorithm and the asterisk$*$ denotes a random perturbation. After that, the adjusted estimates are yielded by resampling the set $\mathcal{P}$, i.e,

$$\begin{aligned}\widehat{\theta}^b_{n,t} &= \widehat{\theta}^b_{n-1} - \gamma_n P^b_n \nabla\ell\left(\widehat{\theta}^b_{n-1}; Z_{t,n}\right),\\ \bar{\theta}^b_{n,t} &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}^b_i,\end{aligned} \tag{0.3}$$

where $\mathcal{P}^b = \left\{P^b_i, i = 1, \ldots, B\right\}$ denotes the resampling collection of $\mathcal{P}$ with $b = 1, \ldots, B$ indicating the bth resampling, and we refer to $\bar{\theta}^b_t$ as the estimate obtained by resampling set $\mathcal{P}^b$.

Algorithm 1: TAB-based Opposite Online Learning

Input: Sequential data $S_t (t = 1, \ldots, T)$
Set the number of bootstraps B
Maximum number of iterations N
Hyperparameter d_0
Output: $\mathcal{Q}_T^{\xi}$

Constructing the loss function $\ell(\theta; Z)$
Generating bootstrap set $\mathcal{P}^b, b = 1, \ldots, B$
Set $\mathcal{Q}_0^{\xi} = 0$
for $t = 1, \ldots, T$ **do**
  Update estimates based on Equation (0.3) using a sample from S_t
  Storage the set E_t and M_t
  Set reward function $W_t^L = M_t$ and $W_t^R = -M_t$
  if $\mathcal{Q}_{t-1}^{\xi} \leq 0$ **then**
    $Z_t^{\xi} = W_t^L$and $\mu_t^{\xi} = d_0$
  else $\{Z_t^{\xi} = W_t^R$and $\mu_t^{\xi} = -d_0\}$
  end if
  Update $\mathcal{Q}_{t-1}$ to $\mathcal{Q}_t$ based on Equation (0.6)
end for
return $\mathcal{Q}_T^{\xi}$

Screening of Elements of Data Set E_t

On the basis of the method of the average SGD via bootstrap, we can obtain a batch of estimates at time point t, which are denoted as $E_t = \left\{\bar{\theta}_t^b, b = 1, \ldots, B\right\}$. Notably, although $\bar{\theta}_t^b$ is related to N, the total iteration number, it is typically assumed that N is ignored on the notation and great enough to make the algorithm convergent.

Although $\bar{\theta}_t^b$ is earned via a bootstrap procedure, the convergence of the algorithm is still promised when the loss function is non-convex or the SGD estimates fall into saddle points, resulting in biased estimates. Therefore, it is a natural idea to eliminate those outliers (such as saddle points) from the set E_t as far as possible. Specifically, after arranging the elements of the set E_t in ascending order, we obtain a new set that is still denoted as $\left\{\bar{\theta}_t^1 \leq \bar{\theta}_t^2 \leq \ldots \leq \bar{\theta}_t^B\right\}$. Then we can select a subset $\left\{\bar{\theta}_t^{(\frac{B}{4}+1)} \leq \bar{\theta}_t^{(\frac{B}{4}+2)} \leq \ldots \leq \bar{\theta}_t^{(\frac{3B}{4})}\right\}$ that is still denoted as E_t. To estimate the parameter θ_0, we take the average of the set $E_t = \left\{\bar{\theta}_t^{(\frac{B}{4}+1)} \leq \bar{\theta}_t^{(\frac{B}{4}+2)} \leq \ldots \leq \bar{\theta}_t^{(\frac{3B}{4})}\right\}$, which is denoted as

$$M_t = \frac{\bar{\theta}_t^{(\frac{B}{4}+1)} + \bar{\theta}_t^{(\frac{B}{4}+2)} + \ldots + \bar{\theta}_t^{(\frac{3B}{4})}}{(\frac{B}{2})}.$$

Opposite Online Learning Procedure

In this section, we will use online estimate M_t sequentially to construct a test statistic.

Firstly, to simulate the TAB process, we refer the uncertain rewards as $W_t^L = M_t$ and $W_t^R = -M_t$, and then we will get a random reward W_t^L when pulling the left (L) arm at time t or its opposite rewards W_t^R when pulling the right (R) arm, which is called TAB-based "opposite online learning".

Additionally, $\sigma^2 = \operatorname{var}\left(W_t^L\right) = \operatorname{var}\left(W_t^R\right)$ can be estimated by $\widehat{\sigma}_t^2 = (\sum_{l=1}^t (M_l - \widehat{M}_t)^2)/(t-1)$, where $\widehat{M}_t \triangleq \frac{\sum_{l=1}^T M_l}{t}$. Moreover, since the expectations of W_t^L and W_t^R are different, the two statistics just conform to the condition of the nonlinear central limit theorem (Chen and Epstein 2022), where expectations are different and variances are identical.

Thus, we can also construct a random reward function governed by a given strategy $\xi = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_T\}$. Specifically, the random reward function is denoted as

$$Z_t^{\xi} = \begin{cases} W_t^L, & \text{if } \vartheta_t = 1 \\ W_t^R, & \text{if } \vartheta_t = 2. \end{cases}$$

For example, if $\vartheta_t = 1$ at time point t, the arm "L" will be pulled with a random reward of W_t^L; Similarly, if $\vartheta_t = 2$, the arm "R" will be pulled with a random reward W_t^R. The test statistic Q_t^{ξ} is

$$\mathcal{Q}_t^{\xi} = \frac{1}{T}\sum_{l=1}^{t} Z_l^{\xi} + \frac{1}{\sqrt{T}}\sum_{l=1}^{t}\frac{Z_l^{\xi} - \mu_l^{\xi}}{\widehat{\sigma}_t},$$

where $\mu_t^{\xi} = I(\vartheta_t = 1)(d_0) + I(\vartheta_t = 2)(-d_0)$ and the strategy ξ is characterized by

$$\vartheta_t = 2 - I\left\{\mathcal{Q}_{t-1}^{\xi} \leq 0\right\}. \tag{0.4}$$

Note that the selection of μ_t^{ξ} is an interactive process, where we will choose the arm L if $\mathcal{Q}_{t-1}^{\xi} \leq 0$ and choose the arm R if $\mathcal{Q}_{t-1}^{\xi} \geq 0$ at the time point t. After T days, our final test statistic can be denoted as

$$\mathcal{Q}_T^{\xi} = \frac{1}{T}\sum_{l=1}^{T} Z_l^{\xi} + \frac{1}{\sqrt{T}}\sum_{l=1}^{T}\frac{Z_l^{\xi} - \mu_l^{\xi}}{\widehat{\sigma}_T}. \tag{0.5}$$

The distribution of $\mathcal{Q}_T^{\xi}$ will be stated in Section 3.

In order to update the statistic $\mathcal{Q}_T^{\xi}$ online, two summary statistics $\left\{\mathcal{S}_1^T, \mathcal{S}_2^T\right\}$ are defined as

$$\mathcal{S}_1^T = \sum_{l=1}^{T} Z_l^{\xi}, \quad \mathcal{S}_2^T = \sum_{l=1}^{T} Z_l^{\xi} - \mu_l^{\xi}.$$

So, $\mathcal{Q}_T^{\xi}$ can be updated by:

$$\mathcal{Q}_T^{\xi} = \frac{1}{T}\left\{S_1^{T-1} + Z_T^{\xi}\right\} + \frac{1}{\sqrt{T}\widehat{\sigma}_T}\left\{S_2^{T-1} + Z_T^{\xi} - \mu_T^{\xi}\right\}. \tag{0.6}$$

In summary, the statistic $\mathcal{Q}_t^{\xi}$ is constructed sequentially as follows: Firstly, the tth batch of samples are observed, and then the bootstrap SGD method yields a estimate set E_t. We proceed to construct the rewards $W_t^L$and W_t^R for the left and right arm based on the set E_t. Subsequently, we obtain Z_t^{ξ} and μ_t^{ξ} by using the strategy $\vartheta_t = 2 - I\left\{\mathcal{Q}_{t-1}^{\xi} \leq 0\right\}$. Meanwhile, we calculate the variance $\widehat{\sigma}_t$ based on historical information and update the statistic from $\mathcal{Q}_{t-1}^{\xi}$ to $\mathcal{Q}_t^{\xi}$.Finally, The details of the opposite online learning method are summarized in Algorithm 1.

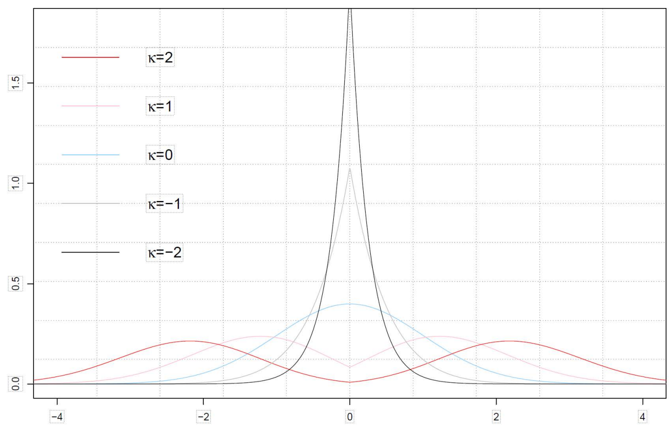

Figure 1: The distribution of $\mathcal{B}(\kappa)$ with different κ.

Theoretical Results

A Novel Distribution

The limit distributions of the statistics $\mathcal{Q}_T^{\xi}$ are several special classes of the following distribution family given by (Chen, Feng, and Zhang 2022).

DEFINITION 1 *We define the density function of limit distribution family, i.e.,*

$$f^{\kappa}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(|y|-\kappa)^2}{2}} - \kappa e^{2\kappa|y|}\Phi(-|y|-\kappa), \quad (0.7)$$

where Φ denotes the distribution function of the standard normal distribution. The variable Y conforms to the above distribution, which is denoted as $Y \sim \mathcal{B}(\kappa)$ with parameter $\kappa \in \mathbb{R}$.

(1) *If $\kappa < 0$, the distribution of Y is sharp and gets sharper with decreasing κ.*

(2) *If $\kappa > 0$, the distribution of Y shows two peaks and gets flatter as κ increases.*

(3) *If $\kappa = 0$, the distribution of Y is standard normal distribution.*

Asymptotic Distribution Under the Null and Alternative Hypotheses

The following theorem gives the limiting distribution of the statistic $\mathcal{Q}_T^{\xi}$ and a generalized form has been proved by (Chen, Feng, and Zhang 2022). When using the estimators produced by SGD, some assumptions need to be satisfied and the details are presented in the Appendix.

THEOREM 1 *Suppose that $\varphi \in C(\overline{\mathbb{R}})$, a set of all continuous functions on $\mathbb{R}$ with finite limits at $\pm\infty$, is a even function and monotone on $(0, \infty)$.*

(1) *For any fixed $d_0, T \geq 1$, we have*

$$\lim_{T\to\infty} \left\{ \mathrm{E}[\varphi(\mathcal{Q}_T^{\xi})] - \mathrm{E}[\varphi(\sigma_{d_0}\eta_T)] \right\} = 0$$

where $\eta_T \sim \mathcal{B}(\kappa_T), \kappa_T = [\sqrt{T}(d_0 - \theta_0)/\sigma] - \theta_0$ and $\sigma_{d_0} = \sqrt{1 + (\theta_0 - d_0)^2/\sigma^2}$.

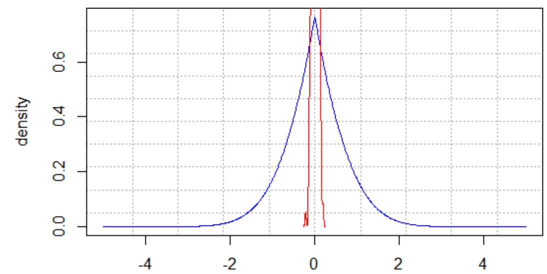

Figure 2: $d_0 = 0.5, \theta_0 = 0.6$

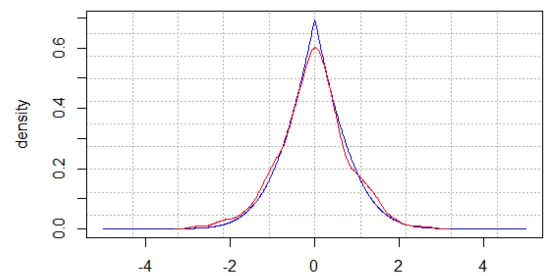

Figure 3: $d_0 = \theta_0 = 0.5$

(2) *For any $y \in \mathbb{R}$, if φ denotes a indicator function on the interval $[-y, y]$, we have*

$$\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi}\right| \leq y\right) = \lim_{T\to\infty} \left[\Phi\left(\kappa_T + \frac{y}{\sigma_{d_0}}\right) - e^{-\frac{2y\kappa_T}{\sigma_{d_0}}} \Phi\left(\kappa_T - \frac{y}{\sigma_{d_0}}\right)\right],$$

where Φ denotes the distribution function of the standard normal distribution.

LEMMA 1 *For any $y \in \mathbb{R}$, suppose that φ denotes a indicator function on the interval $[-y, y]$ and θ_0 is equal to d_0, we will get $\kappa_T = -\theta_0$, i.e., $\mathcal{Q}_T^{\xi} \sim \mathcal{B}(\kappa_T)$ (see Figure 3), inducing*

$$\lim_{T\to\infty} P\left(\left|\mathcal{Q}_T^{\xi}\right| \leq y\right) = \Phi(\theta_0 + y) - e^{-2\theta_0 y}\Phi(\theta_0 - y).$$

REMARK 1 *THEOREM 1 indicates that the limiting distribution of $\mathcal{Q}_T^{\xi}$ is determined by κ_T, which is associated with d_0 and θ_0. Specifically, when $\mathbf{H}_0$ holds, i.e., $d_0 \leq \theta_0$, it means $\kappa_T \leq -d_0$, inducing the distribution of $\mathcal{Q}_T^{\xi}$ will be sharper than the distribution $\mathcal{B}(-d_0)$ (see the red line of Figure 2).*

In the Figures 2-4, the blue lines denote the distribution of η, $\mathcal{B}(-d_0)$, with $d_0 = 0.5$ and the red blue lines are estimated

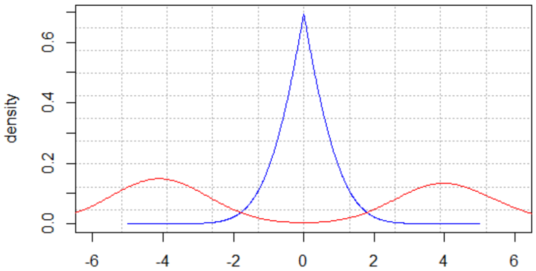

Figure 4: $\theta_0 = 0.48, d_0 = 0.5$

density plots of $\mathcal{Q}_T^{\xi}$ with different θ_0. Moreover, streaming data is generated by the mean model, $Z = \theta_0 + \epsilon$, with $T = 500, B = 50$, in which ϵ denotes a continuous Gaussian noise with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = 1$. The empirical density function is plotted by 1000 replicates $\mathcal{Q}_T^{\xi}$.

REMARK 2 *If T is big enough and $\mathbf{H}_1$ holds, we can also know that $\kappa_T \geq 0$. Hence, the distribution of $\mathcal{Q}_T^{\xi}$ is flat and aggregates at both ends of the plot(see the red line of Figure 4).*

Rejection Region

Invoking THEOREM 1, we can calculate the rejection fields. For any $d_0 > 0$ and $0.5 > \alpha > 0$, the corresponding critical value $z_{\frac{\alpha}{2}}$ is defined by

$$\Pr\left(|Y| < z_{\frac{\alpha}{2}}\right) = 1 - \alpha,$$

where $Y \sim \mathcal{B}(-d_0)$. Following the LEMMA 1, $z_{\frac{\alpha}{2}}$ can be calculated from the following equation:

$$\Phi\left(z_{\frac{\alpha}{2}} + d_0\right) - e^{-2d_0 z_{\frac{\alpha}{2}}} \Phi\left(-z_{\frac{\alpha}{2}} + d_0\right) = 1 - \alpha,$$

where Φ denotes the distribution function of the standard normal distribution.

Since $z_{\frac{\alpha}{2}}$ can be calculated by the theoretical distribution $\mathcal{B}(-d_0)$ with $d_0 = \theta_0$, we assert that the rejection region is

$$(-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, +\infty).$$

Next, we will explain this point. Reviewing the Remarks 1 and 2, we can learn that the null hypothesis $d_0 > \theta_0$ is equivalent to $\kappa_T \leq -d_0$, so that the limit distribution of $\mathcal{Q}_T^{\xi}$ accumulates around 0, i.e.,

$$\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi}\right| < z_{\frac{\alpha}{2}}\right) > 1 - \alpha.$$

Under hypothesis $\mathbf{H}_1$, there is a constant C which makes κ_T greater than 0 when T is bigger than C. Thus, the limit distribution of $\mathcal{Q}_T^{\xi}$ slips to both ends of x-axis, which induces

$$\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi}\right| > z_{\frac{\alpha}{2}}\right) > \alpha,$$

Algorithm 2: An Extended Two-sided Test

Input: Sequential data $S_t (t = 1, \ldots, T)$
Set the number of bootstraps B
Maximum number of iterations N
Hyperparameter $d_0 = 0$

Output: $\mathcal{Q}_T^{\xi_a}, \mathcal{Q}_t^{\xi_b}$

Constructing the loss function $\ell(\theta; Z)$
Generating bootstrap set $\mathcal{P}^b$, $b = 1, \ldots, B$
Set $\mathcal{Q}_0^{\xi_a} = 0$ and $\mathcal{Q}_0^{\xi_b} = 0$
for $t = 1, \ldots, T$ **do**
  Update parameter θ based on Equation (0.3) using data S_t
  Storage estimators E_t and M_t
  Set reward function $W_t^L = M_t$ and $W_t^R = -M_t$
  if $\mathcal{Q}_{t-1}^{\xi_a} \leq 0$ **then**
    $Z_t^{\xi_a} = W_t^L$ and $\mu_t^{\xi_a} = d_0$
  else $\{Z_t^{\xi_a} = W_t^R$ and $\mu_t^{\xi_a} = -d_0\}$
  end if
  Update $\mathcal{Q}_{t-1}^{\xi_a}$ to $\mathcal{Q}_t^{\xi_a}$ based on Equation (0.6)
  if $\mathcal{Q}_{t-1}^{\xi_b} \geq 0$ **then**
    $Z_t^{\xi_b} = W_t^L$ and $\mu_t^{\xi_b} = d_0$
  else $\{Z_t^{\xi_b} = W_t^R$ and $\mu_t^{\xi_b} = -d_0\}$
  end if
  Update $\mathcal{Q}_{t-1}^{\xi_b}$ to $\mathcal{Q}_t^{\xi_b}$
end for
return $Q_T^{\xi_a}$ and $Q_T^{\xi_b}$
If $\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}}$ or $\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}}$, we reject $\mathbf{H_0}$

and the probability, test power, can even approach 1 with the increasing T. To sum up, when $\mathcal{Q}_T^{\xi}$ falls into the rejection field:

$$(-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, +\infty),$$

we reject the null hypothesis $\mathbf{H}_0$.

An Extension to Two-sided Test

Let us consider the following classical two-sided test:

$$\mathbf{H}_0 : \theta_0 = 0; \quad \mathbf{H}_1 : \theta_0 \neq 0.$$

Two-sided test is important for linear regression, because the effect of each predictor variable on the response variable may not always be significant. So we usually perform a significance test for each predictor variable. In this section, we proposed a new significance test procedure based on the TAB model where $\theta_0 = 0$ in regression model implies that the corresponding covariate has no contribution to the response. The above hypothesis can be converted into the following two one-sided hypothesis:

$$\mathbf{H}_{a0} : \theta_0 < 0; \quad \mathbf{H}_{a1} : \theta_0 \geq 0, \tag{0.8}$$

$$\mathbf{H}_{b0} : \theta_0 > 0; \quad \mathbf{H}_{b1} : \theta_0 \leq 0. \tag{0.9}$$

When both hypothesises $\mathbf{H}_{a0}$ and $\mathbf{H}_{b0}$ are rejected, we can accept hypothesis $\mathbf{H}_0$; But if only one of them is accepted,

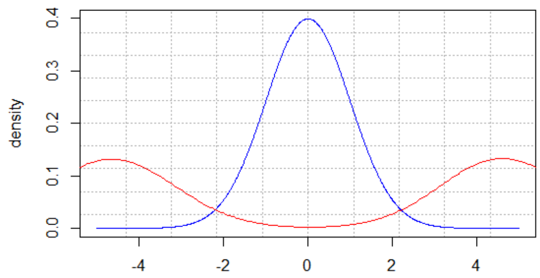

Figure 5: $\theta_0 = 0.02$

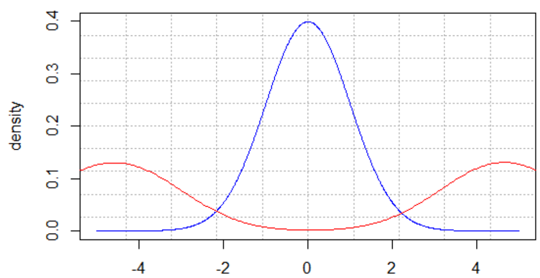

Figure 6: $\theta_0 = -0.02$

the hypothesis $\mathbf{H}_0$ will also be rejected. Then, we can construct a statistic to test the hypothesises (0.8) and (0.9) after applying the theoretical results in Section 3.

Notably, the strategy $\xi_a = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_T\}$ is identical with ξ (see Equation (0.4)) in Section 2. Similarly, the strategy $\xi_b = \{\beta_1, \beta_2, \ldots, \beta_T\}$ can be defined as follows, i.e.,

$$\beta_t = 2 - I\left\{\mathcal{Q}_{t-1}^{\xi_b} \geq 0\right\}.$$

As same as in Section 2, we can construct two reward functions $Z_t^{\xi_a}$ and

$$Z_t^{\xi_b} = \begin{cases} W_t^L, & \text{if } \beta_t = 1. \\ W_t^R, & \text{if } \beta_t = 2. \end{cases} \tag{0.10}$$

Then, the testing statistic are $\mathcal{Q}_T^{\xi_a}$ and

$$\mathcal{Q}_T^{\xi_b} = \frac{1}{T}\sum_{l=1}^{T} Z_l^{\xi_b} + \frac{1}{\sqrt{T}}\sum_{l=1}^{T}\frac{Z_l^{\xi_b} - \mu_l^{\xi_b}}{\widehat{\sigma}_T}, \tag{0.11}$$

where $\mu_t^{\xi_a}$ and $\mu_t^{\xi_b}$ is equal to 0 due to $d_0 = 0$.

LEMMA 2 *For any $y \in \mathbb{R}$, φ denotes a indicator function on the interval $[-y, y]$.*

$$\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_b}\right| \leq y\right)$$
$$= \lim_{T\to\infty}\left[\Phi\left(\kappa_T + \frac{y}{\sigma_{d_0}}\right) - e^{-\frac{-2y\kappa_T}{\sigma_{d_0}}}\Phi\left(\kappa_T - \frac{y}{\sigma_{d_0}}\right)\right],$$

where $\kappa_T = [\sqrt{T}\theta_0/\sigma] + \theta_0$ and $\sigma_{d_0} = \sqrt{1+\theta_0^2/\sigma^2}$ with $d_0 = 0$.

On the one hand, if $\mathbf{H}_0$ holds, the distributions of $\mathcal{Q}_T^{\xi_a}$ and $\mathcal{Q}_T^{\xi_b}$ are $\mathcal{B}(0)$, inducing

$$\begin{aligned} \Pr{}^a &\triangleq \lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| < z_{\frac{\alpha}{2}}\right) = 1-\alpha, \\ \Pr{}^b &\triangleq \lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_b}\right| < z_{\frac{\alpha}{2}}\right) = 1-\alpha, \end{aligned} \tag{0.12}$$

where $z_{\frac{\alpha}{2}}$ is the upper αth quantile of the standard normal distribution. Obviously, under the null hypothesis $\mathbf{H}_0$ we have

$$\begin{aligned} &\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| < z_{\frac{\alpha}{2}} \text{ and } \left|\mathcal{Q}_T^{\xi_b}\right| < z_{\frac{\alpha}{2}}|\right) \\ &= \Pr{}^a + \Pr{}^b - \lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| < z_{\frac{\alpha}{2}} \text{ or } \left|\mathcal{Q}_T^{\xi_b}\right| < z_{\frac{\alpha}{2}}|\right) \\ &> \Pr{}^a + \Pr{}^b - 1 = 1 - 2\alpha, \end{aligned}$$

Therefore, if the conditions, $\left|\mathcal{Q}_T^{\xi_a}\right| < z_{\frac{\alpha}{2}}$ and $\left|\mathcal{Q}_T^{\xi_b}\right| < z_{\frac{\alpha}{2}}$, are satisfied, we will accept $\mathbf{H}_0$ with the probability of making the first type of error being less than 2α.

On the another hand, if $\mathbf{H}_1$ is true, which means $\mathbf{H}_{a0}$ or hypothesis $\mathbf{H}_{b0}$ will be accepted, the limit distribution of $\mathcal{Q}_T^{\xi_a}$ or $\mathcal{Q}_T^{\xi_b}$ will slip to both end according to THEOREM 1 and LEMMA 2. To be more specific, when $\theta_0 > 0$, the distribution of $\mathcal{Q}_T^{\xi_b}$ shows two gentle peaks while the density plot of $\mathcal{Q}_T^{\xi_a}$ is more sharp than the normal density plot. Similarly, when $\theta_0 < 0$, $\mathcal{Q}_T^{\xi_a}$ also shows two gentle peaks while the density plot of $\mathcal{Q}_T^{\xi_b}$ is more sharp than the standard normal density plot. Therefore, when the results, $\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}}$ or $\left|\mathcal{Q}_T^{\xi_b}\right| > z_{\frac{\alpha}{2}}$, is found, we will reject $\mathbf{H}_0$.

After that, we will to seek the probability of making the second type error. If the alternative hypothesis $\mathbf{H}_1$ is right, we have

$$\begin{aligned} &\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}} \text{ or } \left|\mathcal{Q}_T^{\xi_b}\right| > z_{\frac{\alpha}{2}}\right) \\ &> \max\left\{\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}}\right), \lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_b}\right| > z_{\frac{\alpha}{2}}\right)\right\}. \end{aligned}$$

$$\max\left\{\lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_a}\right| > z_{\frac{\alpha}{2}}\right), \lim_{T\to\infty} \Pr\left(\left|\mathcal{Q}_T^{\xi_b}\right| > z_{\frac{\alpha}{2}}\right)\right\}$$

is a relatively large number because the distribution of $\mathcal{Q}_T^{\xi_a}$ or $\mathcal{Q}_T^{\xi_b}$ slides to both ends of the number axis under $\mathbf{H}_1$, which induces that the power is large.

Example 1: If streaming data is generated by a toy model $Z = \theta_0 + \epsilon$ and the noise ϵ is a normal distribution with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = 1$, the distribution of $\mathcal{Q}_T^{\xi_b}$ is displayed in the Figure 5 where the blue line denotes the distribution $\mathcal{B}(0)$ and the red blue line is the empirical density plots of $\mathcal{Q}_T^{\xi_b}$ with $B = 100$, $T = 1000$ after 1000 replicates. Moreover, the simulation shows that the power is as large as 0.9943 with $\theta_0 = 0.02$.

Similarly, if θ_0 is equal to -0.02, the density plots (the red line) of $\mathcal{Q}_T^{\xi_a}$ are is displayed in the Figure 6.As showed in simulation, the power is 0.999 under $\theta_0 = -0.02, \alpha = 0.05, B = 100, T = 1000$.

Finally, the Algorithm 2 for the two-sided test is shown .

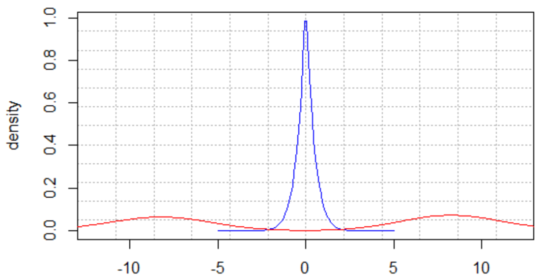

Figure 7: $d_0 = 1.01$ and $\theta_0 = 1$

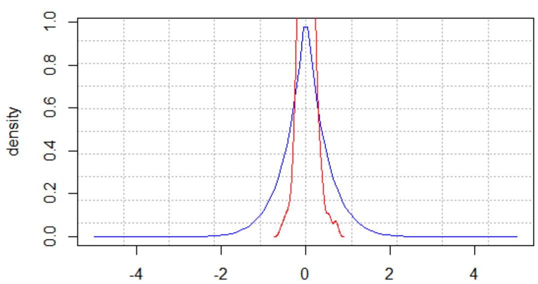

Figure 8: $d_0 = 0.9, \theta_0 = 1$

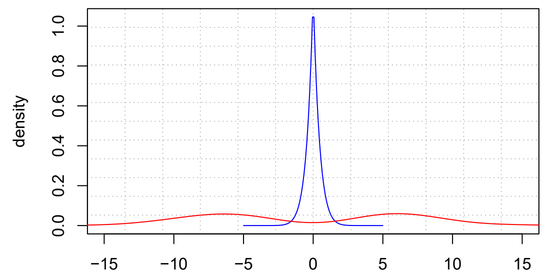

Figure 9: $d_0 = 1.1, \theta_0$=1

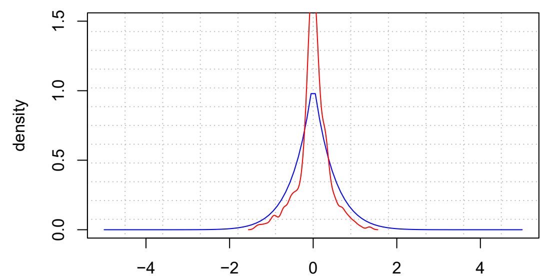

Figure 10: $d_0 = 1, \theta_0 = 1.1$

Simulation

We consider some streaming data is sequentially generated by the following model, i.e.,

$$Y = f(X, \beta) + \epsilon$$

where f is the function to be fitted, and ϵ is a random noise, and β is related to as the parameter θ_0.

After that, we are interested in: (1)whether $\mathcal{Q}_T^{\xi_a}$ fits the theoretical distribution $\mathcal{B}\left(-d_0\right)$; (2)whether the distribution of $\mathcal{Q}_T^{\xi_a}$ is sharper than $\mathcal{B}\left(-d_0\right)$ with $d_0 > \theta_0$; (3)whether the distribution of $\mathcal{Q}_T^{\xi_a}$ is smoother than $\mathcal{B}\left(-d_0\right)$ with $d_0 < \theta_0$.

Example 2: In this example, we consider the one-sided hypothesis testing problem in Section 2 and data is generated by a linear regression model, which means $f(X, \theta_0) = \theta_0 * X$. The loss function is the mean square error and 1-dimensional parameter vector θ_0 is firstly considered. The blue lines in the Figures 8-7 indicate the statistic theoretical distribution and the red lines indicate the simulated distribution. Here, We obtained 300 replicates with $T = 200, N = 100, B = 30$.

When the hypothesis $\mathbf{H}_0$, $d_0 > \theta_0$, holds, the distribution of $\mathcal{Q}_T^{\xi}$ is sharper than the distribution $\mathcal{B}(-d_0)$ (see Figure 8). When the hypothesis $\mathbf{H}_1$, $d_0 > \theta_0$, holds, the distribution of $\mathcal{Q}_T^{\xi}$ slips to two ends (see Figure 7).

Example 3: In this example (Figure 9 and Figure 10), We consider multi-dimensional linear model $f(X, \beta) = \beta^T X$, where $\beta = (0.2, 0.4, 0.6, 0.8, \theta_0)^{\mathrm{T}}$ and the loss function is the mean square error. Aiming to test $\theta_0 \geq d_0$ or $\theta_0 < d_0$, we draw the blue lines, which are the plot of distribution $\mathcal{B}(-1.1)$ and $\mathcal{B}(-1)$. Finally, the red lines are drawn by 300 replicates of $\mathcal{Q}_T^{\xi}$ when $T = 30, N = 500, B = 30, d_0 = 1$ and θ_0 takes different values.

Discussion

The opposite online learning method proposes a test statistic construction that incorporates the “knowledge” from the null hypothesis, which is that the expectation of the left arm under our proposed strategy is greater than 0 when the null hypothesis is true. Therefore, the test statistic incorporating prior knowledge and data contains more information, which is the intuitive understanding of large power enhancement.

Acknowledgements

We are especially grateful to the anonymous reviewers for great feedback on the paper. In addition, Dr. Yan was supported by National Natural Science Foundation of China (Number: 11901352), National Statistical Science Research Project (Number: 2022LY080) and Jinan Science and Technology Bureau (Number: 2021GXRC056). Dr. Kong was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), University of Alberta/Huawei Joint Innovation Collaboration, Huawei Technologies Canada Co., Ltd., and Canada Research Chair in Statistical Learning.

.

References

Bellman, R. 1956. A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4): 221–229.

Bradt, R. N.; Johnson, S.; and Karlin, S. 1956. On sequential designs for maximizing the sum of n observations. *The Annals of Mathematical Statistics*, 27(4): 1060–1074.

Chen, H.; Lu, W.; and Song, R. 2021. Statistical inference for online decision making via stochastic gradient descent. *Journal of the American Statistical Association*, 116(534): 708–719.

Chen, W.; Wang, Y.; and Yuan, Y. 2013. Combinatorial multi-armed bandit: General framework and applications. In *International conference on machine learning*, 151–159. PMLR.

Chen, X.; Lee, J. D.; Tong, X. T.; and Zhang, Y. 2020. Statistical inference for model parameters in stochastic gradient descent. *The Annals of Statistics*, 48(1): 251–273.

Chen, Z.; and Epstein, L. G. 2022. A central limit theorem for sets of probability measures. *Stochastic Processes and their Applications*, 152: 424–451.

Chen, Z.; Feng, S.; and Zhang, G. 2022. Strategy-Driven Limit Theorems Associated Bandit Problems. *arXiv preprint arXiv:2204.04442*.

Chen, Z.; Feng, X.; Liu, S.; and Yan, X. 2023. Optimal distributions of rewards for a two-armed slot machine. *Neurocomputing*, 518: 401–407.

Deshpande, Y.; Javanmard, A.; and Mehrabi, M. 2021. Online debiasing for adaptively collected high-dimensional data with applications to time series analysis. *Journal of the American Statistical Association*, 1–14.

Fan, J.; Gong, W.; Li, C. J.; and Sun, Q. 2018. Statistical sparse online regression: A diffusion approximation perspective. In *International Conference on Artificial Intelligence and Statistics*, 1017–1026. PMLR.

Feller, W. 1991. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons.

Gittins, J.; Glazebrook, K.; and Weber, R. 2011. *Multi-armed bandit allocation indices*. John Wiley & Sons.

Han, R.; Luo, L.; Lin, Y.; and Huang, J. 2021. Online Debiased Lasso for Streaming Data. *arXiv preprint arXiv:2106.05925*.

Jacko, P. 2019. The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths. *arXiv preprint arXiv:1906.10173*.

Kang, S.-H.; and Kim, Y. 2014. Sample size calculations for the development of biosimilar products. *Journal of Biopharmaceutical Statistics*, 24(6): 1215–1224.

Lai, T. L.; Robbins, H.; et al. 1985. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1): 4–22.

Langford, J.; Li, L.; and Zhang, T. 2009. Sparse Online Learning via Truncated Gradient. *Journal of Machine Learning Research*, 10(3).

Lattimore, T.; and Szepesvári, C. 2020. *Bandit algorithms*. Cambridge University Press.

Lin, N.; and Xi, R. 2011. Aggregated estimating equation estimation. *Statistics and its Interface*, 4(1): 73–83.

Liu, R.; Yuan, M.; and Shang, Z. 2022. Online statistical inference for parameters estimation with linear-equality constraints. *Journal of Multivariate Analysis*, 105017.

Luo, L.; and Song, P. X.-K. 2019. Renewable Estimation and Incremental Inference in Generalized Linear Models with Streaming Data Sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1): 69–97.

Mandt, S.; Hoffman, M. D.; and Blei, D. M. 2017. Stochastic gradient descent as approximate bayesian inference. *arXiv preprint arXiv:1704.04289*.

Mel'nikov, A. V. 1979. On strong solutions of stochastic differential equations with nonsmooth coefficients. *Theory of Probability & Its Applications*, 24(1): 147–150.

Nguyen, H.-L.; Woon, Y.-K.; and Ng, W.-K. 2015. A survey on data stream clustering and classification. *Knowledge and information systems*, 45(3): 535–569.

Peng, S. 2008. A new central limit theorem under sublinear expectations. *arXiv preprint arXiv:0803.2656*.

Perchet, V.; and Rigollet, P. 2013. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 41(2): 693–721.

Polyak, B. T.; and Juditsky, A. B. 1992. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4): 838–855.

Ramprasad, P.; Li, Y.; Yang, Z.; Wang, Z.; Sun, W. W.; and Cheng, G. 2022. Online bootstrap inference for policy evaluation in reinforcement learning. *Journal of the American Statistical Association*, 00(0): 1–14.

Robbins, H. 1952. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5): 527–535.

Robbins, H.; and Monro, S. 1951. A Stochastic Approximation Method. *Annals of Mathematical Statistics*, 22(3): 400–407.

Ruppert, D. 1988. Efficient Estimations from a Slowly Convergent Robbins-Monro Process. *Cornell University Operations Research Industrial Engineering*, (781).

Schifano, E. D.; Wu, J.; Wang, C.; Yan, J.; and Chen, M.-H. 2016a. Online updating of statistical inference in the big data setting. *Technometrics*, 58(3): 393–403.

Schifano, E. D.; Wu, J.; Wang, C.; Yan, J.; and Chen, M.-H. 2016b. Online updating of statistical inference in the big data setting. *Technometrics*, 58(3): 393–403.

Shi, C.; Song, R.; Lu, W.; and Li, R. 2021. Statistical inference for high-dimensional models via recursive online-score estimation. *Journal of the American Statistical Association*, 116(535): 1307–1318.

Slivkins, A.; et al. 2019. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2): 1–286.

Sun, L.; and Barbu, A. 2021. A novel framework for online supervised learning with feature selection. *Online Learning*, (1/36).

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.

Thompson, W. R. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4): 285–294.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25(A): 287–298.

Zhu, W.; Chen, X.; and Wu, W. B. 2021. Online covariance matrix estimation in stochastic gradient descent. *Journal of the American Statistical Association*, 1–12.