

Adaptive Entropy Balancing via Multiplicative Weights^{*}

Gaurav Sood[†]

July 16, 2025

Abstract

In applied survey research and causal inference, entropy balancing is used to reweight observational data so that covariate distributions match known or desired population margins. While the standard approach solves a convex optimization problem that minimizes KL divergence from base weights while enforcing exact moment-matching constraints, real-world applications increasingly demand adaptive solutions for streaming data, updated calibration targets, and high-dimensional covariates. We propose replacing closed-form optimization with a Multiplicative Weights Update (MWU) algorithm that maintains weights via iterative exponential updates, operates in batch or streaming mode, and adapts dynamically to changing target moments. Our approach transforms entropy balancing into a no-regret learning process over marginal constraints, offering superior scalability and flexibility while maintaining accuracy comparable to classical methods.

1 Introduction

Entropy balancing has become a fundamental tool in survey research and causal inference for adjusting observational data to match population characteristics (Hainmueller, 2012). The method addresses the common challenge of ensuring that sample covariate distributions align with known population margins, thereby reducing selection bias and improving the validity of statistical inferences.

However, contemporary applications face increasingly complex demands that challenge traditional entropy balancing approaches:

1. **Streaming data requirements:** Modern data collection often involves batched arrivals (e.g., polling waves, rolling panels) that require real-time adjustment rather than batch processing.

^{*}<https://github.com/finite-sample/adaptive-eb>.

[†]Gaurav can be reached at gsood07@gmail.com

2. **Dynamic target adjustment:** Calibration targets frequently change (e.g., revised Census benchmarks, updated population estimates) necessitating rapid recomputation.
3. **High-dimensional scalability:** Rich administrative datasets with thousands of covariates strain traditional optimization approaches.

1.1 Limitations of Current Approaches

The classical entropy balancing method, typically implemented via quasi-Newton methods like BFGS, exhibits several limitations in modern applications:

- **Non-adaptive nature:** Each change in targets or data requires solving the optimization problem from scratch.
- **Scalability constraints:** Performance degrades significantly with high-dimensional covariate spaces ($d > 100$).
- **Batch-only operation:** The method assumes complete dataset availability, precluding streaming applications.

1.2 Our Contribution

We propose an adaptive entropy balancing framework based on Multiplicative Weights Updates (MWU) that addresses these limitations while maintaining the theoretical guarantees of classical entropy balancing. Our approach leverages the connection between entropy balancing and online learning to create a streaming-compatible, scalable alternative that adapts seamlessly to changing conditions.

2 Methodology

2.1 Problem Formulation

Given sample covariates $x_i \in \mathbb{R}^d$ for $i = 1, \dots, n$ and target population moments $\bar{x}_{\text{pop}} \in \mathbb{R}^d$, we seek weights $w \in \Delta^{n-1}$ (the probability simplex) such that:

$$\sum_{i=1}^n w_i x_i = \bar{x}_{\text{pop}}, \quad w_i > 0, \quad \sum_{i=1}^n w_i = 1 \quad (1)$$

Classical entropy balancing solves the optimization problem:

$$\min_w \text{KL}(w \| u) \quad \text{subject to the moment constraints} \quad (\text{EB})$$

where u represents base weights (typically uniform) and $\text{KL}(w \| u) = \sum_{i=1}^n w_i \log(w_i / u_i)$ is the Kullback-Leibler divergence.

2.2 Multiplicative Weights Reformulation

The Lagrangian formulation of problem (EB) with dual variables $\lambda \in \mathbb{R}^d$ yields the closed-form solution:

$$w_i(\lambda) = \frac{u_i e^{-\lambda^\top x_i}}{Z(\lambda)}, \quad Z(\lambda) = \sum_{j=1}^n u_j e^{-\lambda^\top x_j} \quad (2)$$

This exponential form naturally suggests a mirror-descent interpretation. We maintain a weight vector $w^{(t)}$ and apply the multiplicative update:

$$w_i^{(t+1)} \propto w_i^{(t)} \exp\left(-\eta(x_i - \bar{x}_{\text{pop}})^\top g^{(t)}\right) \quad (\text{MWU})$$

where $g^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i - \bar{x}_{\text{pop}}$ represents the current moment error and $\eta > 0$ is the learning rate.

2.3 Algorithm Variants

Our framework supports multiple operational modes:

- **Batch mode:** Updates use the complete dataset X at each iteration.
- **Streaming mode:** Applies exponential updates to mini-batches, cycling through the data over multiple epochs.
- **Adaptive mode:** Continues from current weights when targets change, avoiding restart costs.

Data: Covariates $X \in \mathbb{R}^{n \times d}$, targets $\bar{x}_{\text{pop}}$, learning rate η

Result: Balanced weights w

Initialize $w_i^{(0)} = 1/n$ for all i ;

for $t = 0, 1, \dots, T-1$ **do**

Compute moment error: $g^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i - \bar{x}_{\text{pop}}$;

for $i = 1, \dots, n$ **do**

$w_i^{(t+1)} = w_i^{(t)} \exp\left(-\eta(x_i - \bar{x}_{\text{pop}})^\top g^{(t)}\right)$;

end

Normalize: $w^{(t+1)} \leftarrow w^{(t+1)} / \sum_j w_j^{(t+1)}$;

end

Algorithm 1: Multiplicative Weights Update for Entropy Balancing

2.4 Theoretical Properties

The MWU algorithm provides no-regret guarantees with convergence rate $\tilde{\mathcal{O}}(1/\sqrt{T})$ under standard mirror-descent analysis (Arora et al., 2012). Specifically, the cumulative regret satisfies:

$$\sum_{t=1}^T \langle g^{(t)}, w^{(t)} - w^* \rangle \leq \frac{\log n}{\eta} + \frac{\eta T B^2}{2} \quad (3)$$

where w^* is the optimal entropy balancing solution and B bounds the moment constraint violations.

3 Experimental Results

3.1 Simulation Setup

We evaluate our approach using synthetic datasets with the following configuration:

- Sample size: $n = 3000$
- Covariates: $d = 10$ (standard) and $d = 100$ (high-dimensional)
- Covariate generation: i.i.d. $\mathcal{N}(0, I)$, then scaled
- Population targets: $\bar{x}_{\text{pop}}$ sampled uniformly from $[-0.3, 0.3]^d$

We compare four algorithmic approaches:

1. **BFGS**: Classical entropy balancing via quasi-Newton optimization
2. **Batch MWU**: Full dataset updates at each iteration
3. **Streaming MWU**: Mini-batches of size 50 over 10 epochs
4. **Adaptive MWU**: Target shift after 5 epochs, continued optimization

3.2 Results

Table 1 presents performance comparisons across different scenarios and dimensionalities.

Table 1. Performance comparison of entropy balancing methods

Scenario	L2 Error	Time (s)	Comments
BFGS ($d = 10$)	0.0000	0.008	5 iterations
Batch MWU ($d = 10$)	0.0000	0.004	50 iterations
Streaming MWU	0.0042	0.022	10 passes
Adaptive MWU (shift)	0.0028	0.020	smooth adaptation
BFGS ($d = 100$)	0.0001	1.184	44 iterations
MWU ($d = 100$)	0.0000	0.081	200 iterations

3.3 Key Findings

The experimental results demonstrate several important advantages of the MWU approach:

- **Accuracy preservation:** MWU achieves comparable or superior accuracy to classical BFGS optimization.
- **Scalability gains:** In high-dimensional settings ($d = 100$), MWU provides order-of-magnitude speedups while maintaining accuracy.
- **Adaptive capability:** When targets shift mid-optimization, MWU continues seamlessly without restart penalties.
- **Streaming compatibility:** Mini-batch processing maintains reasonable accuracy with controlled computational overhead.

4 Discussion and Conclusions

4.1 Practical Implications

Our MWU-based entropy balancing framework addresses critical limitations of classical approaches:

- **Flexibility:** Handles target changes without restarting optimization procedures.
- **Scalability:** Runtime complexity scales linearly in both sample size n and dimensionality d , avoiding expensive matrix inversions.
- **Streaming readiness:** Identical exponential updates work effectively in both mini-batch and online settings.
- **Interpretability:** Maintains strictly positive weights with intuitive learning rate control via η .

4.2 Future Directions

Several extensions merit investigation:

- Adaptive learning rate schedules for improved convergence
- Integration with modern automatic differentiation frameworks
- Application to complex survey designs with hierarchical constraints
- Extension to inequality moment constraints

4.3 Conclusion

MWU provides a practical, theoretically grounded alternative to classical entropy balancing when adaptivity, streaming data processing, or high-dimensional covariates are important considerations. The method’s simplicity, combined with strong theoretical guarantees and empirical performance, makes it an attractive tool for modern survey research and causal inference applications.

References

- Arora, S., Hazan, E., and Kale, S. (2012). The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.