Cost-aware Generalized α-investing for Multiple Hypothesis Testing

Thomas Cook, [1] Harsh Vardhan Dubey, [1] Ji Ah Lee, [1] Guangyu Zhu, [2] Tingting Zhao, [3] Patrick Flaherty [1]

[1] Department of Mathematics and Statistics, University of Massachusetts Amherst
[2] Department of Statistics, University of Rhode Island
[3] Department of Information Systems and Analytics, Bryant University

Abstract

We consider the problem of sequential multiple hypothesis testing with nontrivial data collection cost. This problem appears, for example, when conducting biological experiments to identify differentially expressed genes in a disease process. This work builds on the generalized α-investing framework that enables control of the false discovery rate in a sequential testing setting. We make a theoretical analysis of the long term asymptotic behavior of α-wealth which motivates a consideration of sample size in the α-investing decision rule. Using the game theoretic principle of indifference, we construct a decision rule that optimizes the expected return (ERO) of α-wealth and provides an optimal sample size for the test. We show empirical results that a cost-aware ERO decision rule correctly rejects more false null hypotheses than other methods. We extend cost-aware ERO investing to finite-horizon testing which enables the decision rule to hedge against the risk of unproductive tests. Finally, empirical tests on a real data set from a biological experiment show that cost-aware ERO produces actionable decisions as to which tests to conduct and if so at what sample size.

Goal

Our goal is to provide an α-investment rule to simultaneously and optimally allocate our statistical error budget and experimental budget across an unknown, and possibly infinite, number of tests via a game-theoretic framework informed by our prior knowledge.

Problem Setup

- Hypotheses arrive sequentially in a stream. At each step, we must decide whether to reject the current null hypothesis without having access to the number of hypotheses (potentially infinite) or the future p-values, but solely based on the previous decisions.
- A testing procedure provides a sequence of significance levels α_j with decision rule

$$R_j = 1 \text{ if } p_j \le \alpha_j, \text{ else } R_j = 0 \tag{1}$$

- In the online setting, we require that the set significance levels depend only on prior tests.

$$\alpha_j = \alpha_j(R_1, R_2, \ldots, R_{j-1}) \tag{2}$$

- We consider the generalized α-investing framework (Aharoni and Rosset (2014)) where we make use of an α-wealth potential function to bound test levels. If reject the hypothesis, we earn some α-wealth back as a reward for a good investment.
- At each test, we have given α-wealth, $W(j-1)$, which is a statistical error budget. For test j we must determine three quantities, α_j, φ_j, ψ_j.
  - α_j is the significance of the j-th test.
  - φ_j is the cost of the j-th test, we subtract this amount from the α-wealth.
  - ψ_j is the reward we collect if we reject the j-th test.
- In general, we make the following update

$$W(j) = W(j-1) - \varphi_j + R_j \psi_j \tag{3}$$

- α_j, φ_j, and ψ_j are determined by an investment rule $\mathcal{I}_{W(0)}(\{R_1, R_2, \ldots R_{j-1}\})$. Creating an investment rule is non-trivial, but in order to control the marginal false discovery rate ($mFDR_\eta$), we must impose the following constraints.

$$\varphi_j \le W(j-1)$$
$$0 \le \psi_j \le \min\left(\frac{\varphi_j}{\rho_j} + \alpha, \frac{\varphi_j}{\alpha_j} + \alpha - 1\right) \tag{4}$$

- Where ρ_j is the *best power* of test j, which is an upper bound of power across the alternative space. In many cases $\rho_j = 1$, although in biological applications there may be some physical limitation on the bounds of the alternative space, allowing the possibility of $\rho_j < 1$.

Related Work

Many investing rules $\mathcal{I}$ have been proposed for generalized α-investing techniques.

Expected Reward Optimal (ERO)

- A natural consideration is to invest α-wealth in a greedy fashion. Aharoni and Rosset (2014) propose that we optimize the expected reward of each test, since adding wealth back into W provides us with a larger budget for future tests.
- An ERO procedure seeks to maximize $\mathbb{E}(R_j)\psi_j$.
- Aharoni and Rosset (2014) show that an ERO procedure selects an α_j, ψ_j, given a user-specified φ_j, as the solution of

$$\frac{\varphi_j}{\rho_j} = \frac{\varphi_j}{\alpha_j} - 1 \text{ and } \psi_j = \min\left(\frac{\varphi_j}{\rho_j} + \alpha, \frac{\varphi_j}{\alpha_j} + \alpha - 1\right) \tag{5}$$

LORD

- Javanmard and Montanari (2018) propose *Levels Based on Recent Discoveries* (LORD) where an investing rule allocates α-wealth to tests based on the time since a recent, or a collection of recent discoveries. Under certain conditions this method can control the false discovery rate.

SAFFRON

- Ramdas et. al (2018) propose SAFFRON, an generalized α-investing method that adaptively estimates the proportion of true-null hypotheses. This method can be viewed as the online version of the Storey-BH method. This is the first method that adapts to the distribution of the streaming hypotheses.

Methodology

- The generalized α-investing decision rule is augmented to include a notion of dollar-wealth $W_\$(j)$

$$(\varphi_j, \alpha_j, \psi_j, n_j) = \mathcal{I}(W_\alpha(0), W_\$(0))(\{R_1, \ldots, R_{j-1}\}), \tag{6}$$

where n_j is the sample size allocated for testing of the j-th hypothesis.
- A natural update plan for the dollar-wealth is

$$W_\$(0) = B \tag{7}$$
$$W_\$(j) = W_\$(j-1) - c_j n_j, \tag{8}$$

where c_j is the per-sample cost for data to test the j-th hypothesis, and B is the initial dollar-wealth.
- The augmented optimization problem is identical to the one in Aharoni and Rosset (2014) with objective $\max_{\varphi_j, \alpha_j, \psi_j, n_j} \mathbb{E}_\theta(R_j)\psi_j$ and constraint $n_j c_j \le W_\$(j)$.
- The resulting optimization problem has an infinite number of solutions because φ_j is not constrained.
- We then cast the objective function in a Bayesian framework by allowing for the specification of the prior probability of the null hypothesis, $q_j = \Pr[\theta_j \in H_j]$.
- **Game Theoretic Formulation:** Suppose that we have a zero-sum game involving two players: the investigator (Player I) and nature (Player II). Nature, independent of the investigator, chooses to hide $\theta_j \in H_j$ with probability q_j and $\theta_j \notin H_j$ otherwise.
- The investigator selects, φ_j, such that they are indifferent as to whether or not to conduct the experiment,

$$(-\varphi_j + \psi_j) \cdot \left[\alpha_j q_j + \rho_j(1 - q_j)\right] + (-\varphi_j) \cdot \left[(1 - \alpha_j) q_j + (1 - \rho_j)(1 - q_j)\right] = 0 \tag{9}$$

- We briefly present the cost-aware α-investing method in algorithmic form.

Algorithm 1: Cost-Aware ERO Algorithm

Input: α, $W_\alpha(0)$, $W_\$(0)$
$j \leftarrow 0$;
while $W_\alpha(j) > \epsilon$ *and* $W_\$(j) > \epsilon$ **do**
  Increment $j \leftarrow j + 1$;
  Define q_j, c_j for hypothesis j ;
  Solve maximization problem to obtain φ_j, α_j, ψ_j, and n_j ;
  Collect data $(y_{j1}, \ldots, y_{jn_j})$ and compute p-value p_j ;
  if $p_j \le \alpha_j$ **then**
    $R_j \leftarrow 1$
  else
    $R_j \leftarrow 0$
  end
  Update $W_\$(j) \leftarrow W_\$(j) - c_j n_j$;
  Update $W_\alpha(j) = W_\alpha(j) - \varphi_j + R_j \psi_j$;
end

FDR and mFDR

- $R(m) = \sum_{j=1}^m R_j$ and $V(m) = \sum_{j=1}^m V_j$, where $V_j \in \{0, 1\}$ indicates whether the test H_j is both true and rejected
- With these variable definitions, the FDR is

$$\mathrm{FDR}(m) = P_\theta(R(m) > 0)\, \mathbb{E}_\theta\left[\frac{V(m)}{R(m)} \mid R(m) > 0\right] = \mathbb{E}_\theta\left[\frac{V(m)}{R(m) \vee 1}\right],$$

- And the marginal false discovery rate is

$$\mathrm{mFDR}_\eta(m) = \frac{\mathbb{E}_\theta\left[V(m)\right]}{\mathbb{E}_\theta\left[R(m) + \eta\right]}.$$

Comparison to state of the art methods

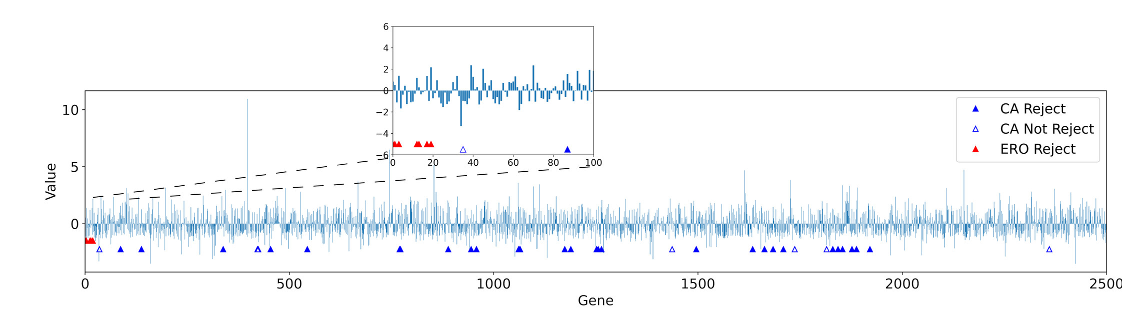
- Simulated 1000 sequential Neyman-Pearson type tests, with 10000 repetitions.
- $H_0 : \theta_j = 0$, $H_1 : \theta_j = 2$, where $\sigma = 1$.
- Under H_0, $Z_j \sim N(0, 1/\sqrt{n})$, under H_1, $Z_j \sim N(2, 1/\sqrt{n})$.
- Each H_j is null with probability 0.9.
- We limit each run to use up to 1000 total samples.
- For non cost-aware schemes, we allow 1 sample per test.

The table is indexed by the φ allocation scheme (Scheme), and the reward method (Method). Together, the Scheme and Method make an investment rule $\mathcal{I}$.

Scheme	Method	Tests	True Rej	False Rej	mFDR
constant	α-spending	10.0	0.27	0.04	0.034
	α-investing	15.9	0.43	0.07	0.046
	α-rewards $k = 1$	15.2	0.42	0.06	0.045
	α-rewards $k = 1.1$	18.5	0.45	0.06	0.042
	ERO investing	18.2	0.49	0.08	0.050
relative	α-spending	66.0	0.54	0.04	0.028
	α-investing	81.1	0.85	0.09	0.047
	α-rewards $k = 1$	80.7	0.83	0.09	0.047
	α-rewards $k = 1.1$	89.8	0.86	0.08	0.043
	ERO investing	82.3	0.89	0.10	0.051
other	LORD++	971.0	2.86	0.08	0.020
	LORD1	1000.0	1.46	0.04	0.018
	LORD2	1000.0	2.02	0.08	0.026
	LORD3	1000.0	2.49	0.08	0.023
	SAFFRON	1000.0	1.57	0.09	0.034
cost-aware	ERO $n_j = 1$	**364.2**	**4.13**	**0.22**	**0.041**
cost-aware	ERO $n_j \le 10$	39.5	3.93	0.20	0.040
cost-aware	ERO $n_j \le 100$	10.8	1.08	0.06	0.026
cost-aware	ERO n_j^*	6.0	0.61	0.03	0.019

Prostate Cancer Gene Expression Data

- Data collection and preprocessing
  - Gene expression data was collected to investigate the molecular determinant of prostate cancer. The data set contains 50 normal samples and 52 tumor samples and each sample is a $m = 6033$ vector of gene expression levels.
  - Considered one-sided Gaussian tests where $\bar{\theta}_j = \log_{10}(2)/\hat{\sigma}_j$.
  - A logistic function, using only the first two samples for each gene, was used to compute the prior probability of the null hypothesis.
  - The set of genes was permuted randomly and the cost-aware and original ERO decision functions were computed.
  - For the ERO comparison, we allocate the maximum number of available samples, $n = 50$, for each test.
  - For cost-aware ERO, if the optimal sample size was greater than the number of available samples ($\bar{n}_j = 50$), the test was skipped, otherwise the one-sided Gaussian test was performed with the optimal number of samples.
  - We set the cost of each sample $c_j = 1$, and $W_\$ = 1000$. Testing concludes when either W_α or $W_\$$ is completely spent.
- Comparison to other algorithms



  - ERO selects many tests, but rapidly expends $W_\$$.
  - Cost-aware ERO is more conservative and only tests when the benefits outweigh the risk of a dual-currency wealth state (W_α, $W_\$$).
  - Across 1000 permutations, cost-aware ERO performed 4.6 tests and skipped 236.8 tests on average.
  - The average optimal sample size was 44.2.

Discussion

Summary

- We extend generalized α-investing to address the problem of online FDR control where the cost of data is not negligible.
- We propose a generalized α-investing procedure for sequential testing that optimizes sample size and φ using the game-theoretic indifference principle.

Limitations

- First, the optimality of the cost-aware ERO method may be sensitive to misspecification in q_j, although $mFDR_\eta$ is still controlled.
- Simulations with an increasing distance between the true q_j and that used by the cost-aware ERO method show that cost-aware ERO performance degrades for misspecified values of q_j
- Cost-aware ERO may aggressively spend available (W_α, $W_\$$) when only considering the reward of a single test.

Future Work

- Investigate a principled risk-hedging approach to conserve some wealth for future tests with the hope that a test with a more favorable reward structure is over the horizon.

References

- Aharoni, E. and Rosset, S. (2014). Generalized α-investing: definitions, optimality results and application to public databases. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(4):771–794.
- Foster, D. P. and Stine, R. A. (2008). α-investing: a procedure for sequential control of expected false discoveries. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(2):429–444.
- Javanmard, A. and Montanari, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. The Annals of statistics, 46(2):526–554.
- Ramdas, A., Zrnic, T., Wainwright, M., and Jordan, M. (2018). Saffron: an adaptive algorithm for online control of the false discovery rate. In International conference on machine learning, pages 4286–4294. PMLR.

Acknowledgments

This project was supported by National Science Foundation HDR TRIPODS 1934846.

UMASS AMHERST

    Email: tjcook@umass.edu, hdubey@umass.edu