A zero-sum game is usually defined by a payoff matrix \( M \). When player 1 chooses action \( i \in \{1, 2, \ldots, I\} \) and player 2 chooses action \( j \in \{1, 2, \ldots, J\} \), then player 1 gets reward \( M_{ij} \) and player 2 gets reward \(-M_{ij}\). Mixed strategies \( x \) and \( y \) for player 1 and 2 are probability distributions on their pure strategies. A best response \( x \) of player 1 to the mixed strategy \( y \) of player 2 is a strategy which maximizes the expected payoff \( Mx \) of player 1. Similarly, a best response of player 2 maximizes his/her expected gain.

A Nash equilibrium is a pair of mixed strategies where neither of the players expects a better reward if he changes unilaterally his strategy. Our goal is to find sparse mixed strategies for two players, and to do it efficiently. We propose an approach which finds compact Nash equilibria. We consider an \( \epsilon \)-Nash equilibrium which is defined as a pair \((x, y)\)

\[
\inf_{x'} Mx' \geq (1-\epsilon) \inf_x Mx - \epsilon,
\]

\[
sup_{y'} My' \leq (1-\epsilon) \sup_y My + \epsilon.
\]

The intuition behind the novel approach is based on the sparsity elimination of those strategies, whose estimated probability values are not significant after the number of rounds \( t \):

\[
x'_i = \begin{cases} 1 & \text{if } x_i > \frac{1}{t}, \\ 0 & \text{if } x_i < \frac{1}{t}, \end{cases}
\]

where \( x_i \) is a coordinate of the probability distribution, \( x'_i \) is a corresponding (sparse) approximation, and parameter \( \epsilon \in (0, 1) \). Note that \( x' \) is associated with a subset of features, and not with one feature.

The probability distribution \( x \) can be estimated by a bandit algorithm such as INF (implicitly normalized predecessors) or EXP3 (exponential exploration-exploitation) algorithms.

### Results

![Experimental Results](image)

- **Figure**: Experiments on the MicroObes transcriptomic data. On the left: accuracy and similarity on the level of separate genes. On the right: stability and functional stability.

- **It is interesting to see that the proposed stochastic thresholding bandit has a very poor similarity and stability values, since the features are randomly drawn. However, the functional stability of the thresholding bandit is higher compared to the state-of-the-art methods, since informative feature sets have high and similar rankings.**

### MicroObes Data

The MicroObes corpus contains meta-data (clinical data), gene expressions of adipose tissue, and gut flora metagenomic data for 49 patients (Pitié-Salpêtrière hospital, Paris, France). For each patient, we have the information to which class he or she belongs: high gene count (HGC) and low gene count (LGC) classes, based on quantitative metagenomics.

### Conclusions

We have introduced a feature selection approach based on a sparse stochastic bandit, and we compared its predictive performance and stability to several state-of-the-art methods which are reported to be the most accurate and stable. We considered the newly introduced feature selection approach from cross-validation error rate and stability viewpoints, and our experiments on artificial data illustrate that the proposed approach is competitive.

We also hope that functional classes of features learned from experiments, will provide additional knowledge and give rise to new hypotheses for biologists and physicians doing pre-clinical research.

### References


### Summary

In large-scale systems biology applications, features are structured in hidden functional categories, whose predictive power is identical. Feature selection, therefore, can load not only to a problem with a reduced dimensionality, but also reveal some knowledge on functional classes of variables. In this contribution, we propose a framework based on a sparse zero-sum game which performs a stable functional feature selection. In particular, the approach is based on feature subsets ranking by a thresholding stochastic bandit. We illustrate by experiments on real complex data that the proposed method is competitive from the predictive and stability viewpoints.

### Motivation

Feature selection is a problem which arises naturally in a number of applications, and, in particular, in biomedical tasks, where the number of parameters is potentially very high but just a small subset of them is informative. It has been observed that functions captured by different feature sets can be very similar, despite a very low degree of overlapping between these feature sets. Our research is motivated by real high-dimensional biomedical applications, in particular, by challenges of quantitative metabolomics and transcriptomics. In quantitative metabolomics we study the collective genome of the micro-organisms inhabiting human body, and, since recently, it has become feasible to measure the abundance of bacterial species. Scientists doing pre-clinical research are often interested to find groups of bacterial species associated with a particular phenomenon.

### Contribution

We propose to apply sparse zero-sum games to select the most pertinent subsets of parameters, and to reach the stability on the functional level. We introduce a thresholding stochastic bandit which efficiently ranks sets of features.

### Stability and Similarity Measures

To compare the outcomes of different feature selection methods, we adopt two measures. The first measure we apply is a similarity measure. The overlap similarity measure between two sets of features \( J_1 \) and \( J_2 \) is computed as the number of variables that are present in both sets, \( |J_1 \cap J_2| \). The second measure we test, is a stability measure \( \alpha \) called relative weighted consistency \( CW(\alpha) \). This measure is subset-size-ubased. Let us carry out feature selection \( \alpha \) times and obtain \( S = \{S_1, \ldots, S_\alpha\} \) subsets of selected features from \( \alpha \) experiments. \( \sum_{\alpha=1}^{\alpha} |S| \) is the sum of cardinalities of sub-sets \( S \) and \( F \) is the number of times the variable \( j \in F \) was observed in \( S \). The CW and \( CW(\alpha) \) measures are defined as follows

\[
 CW(S) = \frac{1}{F} \sum_{j=1}^{F} \frac{1}{\alpha} \sum_{k=1}^{\alpha} \mathbb{1}_{j \in S_k},
\]

\[
 CW(\alpha) = \frac{CW(S) - CW(S_\alpha)}{CW(S_\alpha) - CW(S_\beta)}.
\]