How to group users even if they have very different statistical distribution?

We can think of using either statistical techniques or machine learning to solve the issue. But many standard statistical methods of distributions assume a lot and therefore severely limit the kinds of differences we are looking for. Maybe a supervised classification algorithm in machine learning might come to the rescue?

It might be easier to think about some of these issues if you construct a two-sample test out of a classifier. The procedure is as follows:

* Split your observations X and Y into two parts each, Xtrain and Xtest, Ytrain and Ytest.
* Train a classifier to distinguish between Xtrain and Ytrain.
* Apply the output of the classifier to Xtest and Ytest.
* Count up the portion of times its prediction was correct to get 𝑝̂. Apply a binomial test to distinguish the null p=1/2 from p≠1/2. If p≠1/2, then the two distributions are different.

By inspecting the learned classifier, you may also be able to interpret the differences between the distributions in a semi-meaningful way. By changing the family of classifiers, you consider, you can also help guide the test to look for certain kinds of differences.