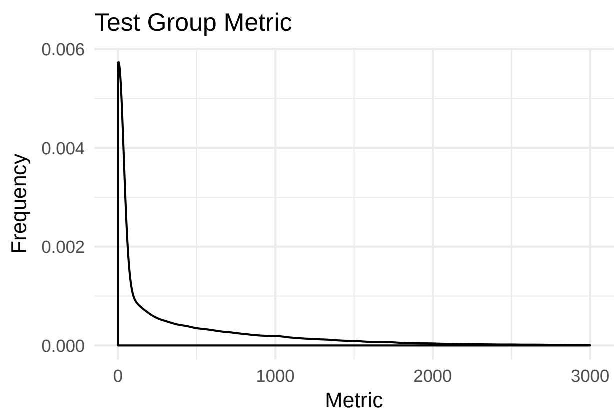
Bootstrapping Non-Normal Data for Significance test

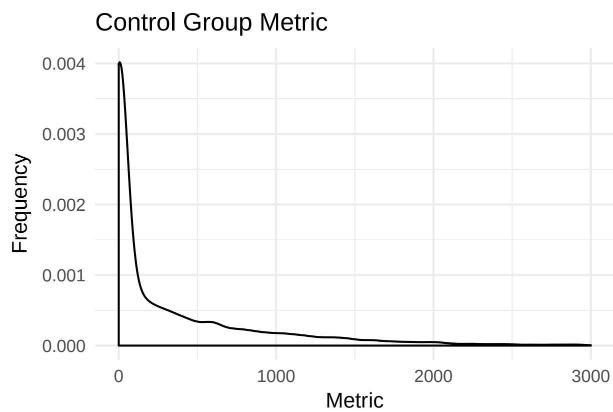
Tawsif Khan

December 17, 2017

A core component of a PCA is comparing the test and control groups on a certain metric and determine whether a campaign has created a statistically significant improvement or not.

Consider the non-binomial metrics like spend data and number of transactions. **These kind of data usually have a skewed distribution and are non-normal.**





Comparing these two samples is difficult because it either requires

- a transformation or,
- tests that don't have normality assumptions.

And, in either case the interpretation of the results are not intuitive. For example a power transformation [1] will raise the distribution to a certain power to make it normal but makes it difficult to communicate the resulting data in business terms. Similarly, a significance test for non-normal data is the two samples Wilcox test [2],[3] which compares the medians of the samples and a location shift of the groups. This is not useful for campaign lift calculations.

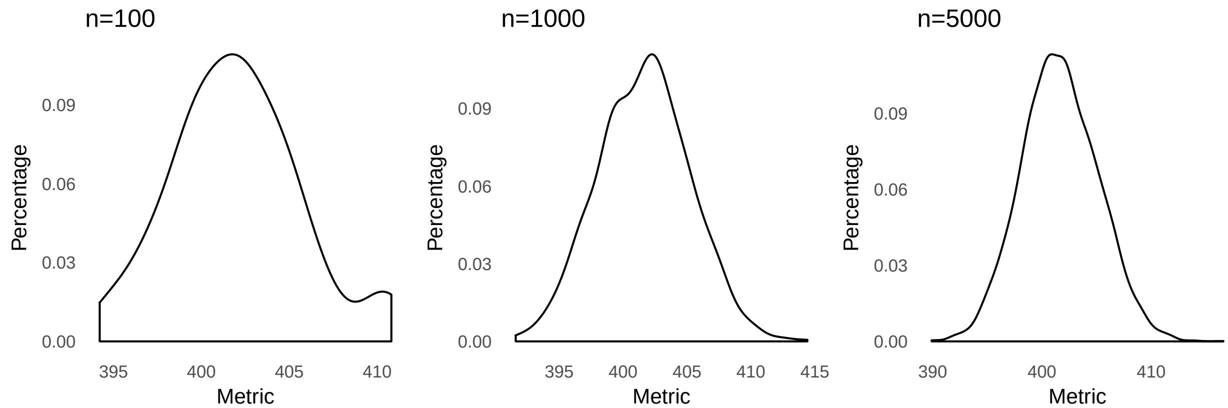
Hence it will be convinient to work with a test that doesn't require changing the domain of the data and allows us to compare metrics of our interest.

Bootstrapping

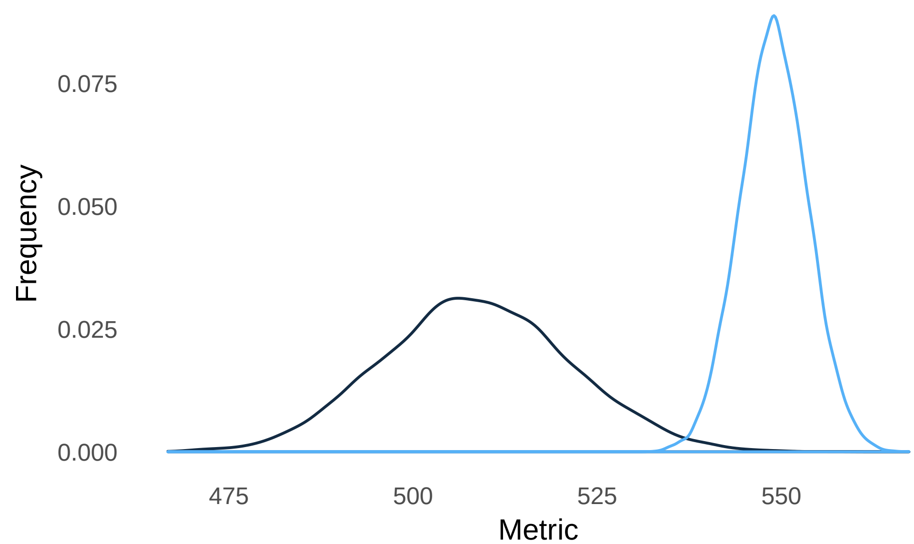
Bootstrapping [4], [5], [6] is a method where the data groups would be repeatedly sampled to estimate the actual means of the populations. It is based on the *central limit theorem* -

- The means of randomly drawn samples will approach the actual mean of the population
- The means of the sample will follow a normal distribution

Distribution of Sample Mean Metric



The figure above showes that after a certain number of iterations the mean metric approaches a normal distribution and the average of these means would be the estimate of the actual populations which we can compare.




```
## [1] "The estimated mean metric of the control group is: 508.584393100532"
```

```
## [1] "The estimated mean metric of the test group is: 549.013581970616"
```

The confidence interval of the mean difference is,

```
quantile(mean_diff_bs,c(.05,.95))
```

```
##      5%     95%
## 18.1024 62.7355
```

Since the confidence interval does not contain 0 (zero), the difference in the means are significant.

Note that, your relative population sizes (control and test) will affect the width of the confidence interval. The larger your sample is the smaller your standard error will be which in turn will return smaller confidence intervals.

Compare Bootstrapping method and Wilcoxon Test

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  metric by flag
## W = 20300000, p-value = 0.04964
## alternative hypothesis: true location shift is not equal to 0
```

Since p-value $<$ .05, the test and control group are signficantly different in Wilcox test as well.

Conclusion

Both Mann-Whitney/Wilcoxon and bootstrapping are examples of permutation based tests. The advantage of bootsrapping however is you can specify a meaningful test statistic.

References

1. Making data normal using box-cox power transformation. http://bit.ly/289Hatf.
2. Mann-whitney-wilcoxon test. http://bit.ly/1J4vEGi.
3. Mann–Whitney u test. http://bit.ly/1LxAuWM.
4. Determination of confidence intervals in non-normal data. http://bit.ly/2r3yBnR.
5. A comparison between normal and non-normal data in bootstrapa. http://bit.ly/2AUyj1U.
6. Bootstrapping - wikipedia. http://bit.ly/1hKKIfM.