**Data 8 Spring 2020**

**Discussion: A/B Testing**

One special kind of hypothesis test we do in this class is called an A/B test. The steps used to run an A/B test are the same as a general hypothesis test, but A/B tests have a specific null hypothesis (that two samples were drawn from the same distribution), which we test by performing a *permutation*.

1.Choose True/False for each of the statements below, and explain your answer.

1. A/B testing is used to determine whether or not we believe two samples come from the same underlying distribution.

True, this is the definition of A/B testing.

1. To conduct a permutation test, you should sample your data with replacement with a sample size equal to the number of rows in the table.

False, you should sample your data without replacement--otherwise, you would not get a permutation of your data.

1. A/B testing is the same as using total variation distance as a test statistic for a hypothesis test.

False, total variation distance is just a test statistic which computes the distance between two categorical distributions. It does not involve taking a random permutation of your data.

1. A/B testing is about comparing two samples, whereas regular hypothesis testing is about testing properties of one sample.

True, A/B testing involves test statistics using two separate samples (like difference between average baby weight of smokers and non-smokers), whereas regular hypothesis testing involves test statistics using just one sample (e.g. absolute difference between expected and actual proportion of heads when flipping a coin 10 times).

2. Natalia, a museum curator, has recently been brought specimens of caddisflies collected from various parts of Northern California. The scientists who collected the caddisflies think that caddisflies collected at higher altitudes tend to be bigger. They tell her that the average length of the 560 caddisflies collected at high elevation is 14mm, while the average length of the 450 caddisflies collected slightly lower down is 12mm. She’s not sure that this difference really matters, and thinks that this could just be the result of chance in sampling.

1. How could you test the null hypothesis in the A/B test from above? What assumption would you make to test the hypothesis, and how would you simulate under that assumption?

If the null hypothesis is true - the caddisflies did not come from different distributions - then it shouldn’t matter how the samples were labeled (high elevation or low elevation). Under this assumption, you could shuffle the labels of the caddisflies and calculate your test statistic from this “relabeled” data.

1. What would be a useful test statistic for the A/B test?

Difference in mean lengths between the two groups. Note that this is not an absolute difference-- we could choose either order for subtraction, but that would affect the direction of our alternative hypothesis so we need to be careful!

1. Assume original\_table refers to the following table:

|  |  |
| --- | --- |
| **A/B label** | **Specimen length (mm)** |
| High elevation | 12.3 |
| Low elevation | 13.1 |
| High elevation | 12.0 |

...

(1007 rows omitted)

Fill in the blanks in this code to generate one value of the test statistic under the null hypothesis.

def permutation\_test():

shuffled\_labels = original\_table.sample(with\_replacement = False).column(“A/B label”)

Original\_with\_shuffled\_labels = original\_table.drop(‘A/B label’).with\_columns(‘A/B label’, shuffled\_labels)

grouped = original\_with\_shuffled\_labels.group(‘A/B label’, np.mean)

means = grouped.column(‘specimen length mean’)

statistic = means.item(0) - means.item(1)

return statistic

1. Fill in the code below to simulate 10000 trials of our permutation test

test\_stats = make\_array()

repetitions = 10000

for i in np.arange(repetitions):

one\_stat = permutation\_test()

test\_stats = np.append(test\_stats, one\_stat)

test\_stats

1. Given that the p-value of the test was 0.1, draw a possible histogram of the test statistics

Histogram centered at 0, about 10% of area > 2 on right hand side (or left hand side depending on what direction our alternative is!)