Interview Questions: Statistical Inference

Signal Data Science

- How would you explain an A/B test to an engineer with no statistics background?
- How would you explain the meaning of a X% confidence interval to an engineer with no statistics background?
  - Repeat process over and over again; X% of the measurements will fall within the confidence interval.
- What is a p-value?
  - Probability of observing the data if the null hypothesis were true.
- In an A/B test, how can you check if assignment to the various buckets was truly random?
- What might be the benefits of running an A/A test, where you have two buckets who are exposed to the exact same product?
  - You can make sure that your A/B testing infrastructure works properly.
- What would be the hazards of letting users sneak a peek at the other bucket in an A/B test?
- What would be some issues if blogs decide to cover one of your experimental groups?
- How would you conduct an A/B test on an opt-in feature?
- How would you run an A/B test for many variants, say 20 or more?
- How would you run an A/B test if the observations are extremely right-skewed?
- I have two different experiments that both change the sign-up button to my website. I want to test them at the same time. What kinds of things should I keep in mind?
- You are AirBnB and you want to test the hypothesis that a greater number of photographs increases the chances that a buyer selects the listing. How would you test this hypothesis?

- How would you design an experiment to determine the impact of latency on user engagement?
- What is maximum likelihood estimation? Could there be any case where it doesn't exist?
  - MLE is estimation of model parameters such that the likelihood of observing the training data is maximized. The maximum likelihood can be infinity. For example, in a Gaussian mixture model, if the center of a Gaussian sits directly on top of a data point, the likelihood can be driven arbitrarily high by making the width of the Gaussian arbitrarily small.
- What's the difference between a MAP, MOM, MLE estimator? In which cases would you want to use each?
- What is a confidence interval and how do you interpret it?
- What is unbiasedness as a property of an estimator? Is this always a desirable property when performing inference? What about in data analysis or predictive modeling?