STAT 450: Statistical Theory

The basics of statistical inference

Definition: A **model** is a family $\{P_\theta; \theta \in \Theta\}$ of possible distributions for some random variable X. (Our data set is X, so X will generally be a big vector or matrix or even more complicated object.)

We will assume throughout this course that the true distribution P of X is in fact some P_{θ_0} for some $\theta_0 \in \Theta$. We call θ_0 the true value of the parameter. Notice that this assumption will be wrong; we hope it is not wrong in an important way. If we are very worried that it is wrong we enlarge our model putting in more distributions and making Θ bigger.

Our goal is to observe the value of X and then guess θ_0 or some property of θ_0. We will consider the following classic mathematical versions of this:

1. Point estimation: we must compute an estimate $\hat\theta = \hat\theta(X)$ which lies in Θ (or something close to Θ).

2. Point estimation of a function of θ: we must compute an estimate $\hat\phi = \hat\phi(X)$ of $\phi = g(\theta)$.

3. Interval (or set) estimation. We must compute a set $C = C(X)$ in Θ which we think will contain θ_0.

4. Hypothesis testing: We must choose between $\theta_0 \in \Theta_0$ and $\theta_0 \notin \Theta_0$ where $\Theta_0 \subset \Theta$.

5. Prediction: we must guess the value of an observable random variable Y whose distribution depends on θ_0. Typically Y is the value of the variable X in a repetition of the experiment.

There are several schools of statistical thinking. The main schools of thought summarized roughly as follows:

- **Neyman Pearson**: A statistical procedure is evaluated by its long run frequency performance. Imagine repeating the data collection exercise many times, independently. Quality of procedure measured by its average performance when true distribution of X values is P_{θ_0}.

- **Bayes**: Treat θ as random just like X. Compute conditional law of unknown quantities given knowns. In particular ask how procedure will work on the data we actually got – no averaging over data we might have got.

- **Likelihood**: Try to combine previous 2 by looking only at actual data while trying to avoid treating θ as random.

This is largely a frequency theory course. We use the Neyman Pearson approach to evaluate the quality of likelihood and other methods. In general statistical theorists work on the following kinds of problems:

- Someone else has suggested a method of analyzing data. Theorists work to evaluate, mathematically, how well this method can be expected to work under a variety of assumptions about the real world – that is, assumimg a variety of models for the real world. For instance, I once wrote a paper called "Overweight tails are inefficient" in which I studied some previously suggested methods of *goodness-of-fit* testing and showed that they had some poor properties in large samples. Sometimes we compare several methods to discover the circumstances in which one would be beter than the others.

- Theory is used to motivate, suggest or derive methods of analyzing data. Likelihood methods, optimal testing theory, unbiased estimating equations all have this flavour.

- Theorists suggest probability models for specific data types and describe appropriate likelihood functions.

- Sometimes we try to find the best possible method of data analysis within some class of possible methods, or to describe principles which allow us to narrow our search for good, or best, methods.