

Lecture 6: Model reliability

1. The basics

- Suppose we have fit a model to a set of data. In doing so, we have determined values for the free parameters of the model. It is useful to know how certain we can be with respect to the specific parameter values that we have obtained.
- The reason that there is uncertainty at all has to do with the distinction between the sample and the population. We are not intrinsically interested in the model parameters that best describe a given sample of data, but rather we are interested in the model parameters that best describe the population from which the sample is drawn.
- A simple analogy may be helpful: Due to sampling error, we can expect some variability in the mean of sample drawn from a distribution and we therefore put error bars on the mean of a given sample of data. Likewise, due to sampling error, we can expect some variability in the model parameters that describe a sample of data. We therefore put error bars on model parameters estimated from a given sample of data.
- We define *model reliability* as the stability, with respect to sampling error, of the parameters of a fitted model. The more data we collect, the more reliable the model parameters will be.

2. Error bars on parameter estimates

- To calculate error bars on parameter estimates, we once again turn to the nonparametric technique of *bootstrapping*. The idea is simple: draw bootstrap samples from the observed data, fit the model of interest to each bootstrap sample, and examine the distribution of parameter estimates across the various bootstrap iterations. Note that this is the same logic we used to put error bars on correlation values (see Lecture 2).
- Note that depending on our goals, we might be interested in something other than the individual model parameters. For example, we might be interested in some metric computed from several parameters or we might be interested in the overall function described by a given model. Bootstrapping handles these scenarios nicely: compute the quantity of interest at each bootstrap iteration, and then look at the distribution of this quantity across the various iterations.
- *Jackknifing* is another resampling technique that can be used to determine sampling error. It is less computationally intensive than bootstrapping but is also less accurate and less flexible. *Split-half analysis* (in which a given set of data is split into halves and each half is analyzed separately) is a simple technique that provides some basic information regarding sampling error. There is no compelling reason to use these alternative techniques unless computational time makes bootstrapping impractical.

3. Parameter estimates may have correlated errors

- For models with more than one parameter, the sampling error is a multidimensional distribution. Thus, putting an error bar on individual parameters may not provide a complete characterization of the sampling error (since there may be interactions between parameters).
- Let's take a simple example: Suppose we have a linear regression model with two regressors that are highly correlated. In this case, an increase in the weight associated with one regressor

can be counteracted by a decrease in the weight associated with the other regressor. So, in a sense, it is easy for the weights to vary in an anti-correlated fashion. Examining variability in weights estimated from different samples, we will find that the error on one weight negatively correlates with the error on the other weight.

- A useful way to think about correlated errors is in terms of the error surface. If the error surface is shallow along certain directions in the parameter space, then errors in the parameter estimates will tend to be correlated along these same directions.

- The weights estimated for correlated regressors will tend to have low reliability. However, even though the individual weights are unreliable, the overall model can still be reliable. That is, given estimates of the individual weights, we can construct the model fit as a weighted sum of the regressors (using the estimated weights), and this model fit can still have high stability (across samples from the population).

4. The distinction between accuracy and reliability

- Accuracy (as discussed in Lecture 5) and reliability (as discussed in the current lecture) are related but are not the same. Accuracy corresponds to how far a fitted model is away from the true model, whereas reliability corresponds to how stable the fitted model is.

- To understand the distinction between accuracy and reliability, let us examine four possible scenarios.

- *Scenario 1: Model is reliable and accurate.* In this case, there are enough data to obtain stable estimates of the parameters of the model (the parameters do not vary much from sample to sample), and the fitted model is close to the true model.

- *Scenario 2: Model is reliable but inaccurate.* In this case, there are enough data to obtain stable estimates of the parameters of the model, but the fitted model is far from the true model.

- *Scenario 3: Model is unreliable but accurate.* This is not a likely situation. Suppose there are insufficient data to obtain stable estimates of the parameters of the model. This implies that the predictions of the model would fluctuate widely from sample to sample. Thus, it is unlikely that the model fitted to any given sample will be close to the true model.

- *Scenario 4: Model is unreliable and inaccurate.* In this case, there are insufficient data to obtain stable estimates of the parameters of the model, and the fitted model is far from the true model. One explanation for why the fitted model is far from the true model is that there is simply a lack of sufficient data to fit the model. An alternative explanation is that the model being used is an inappropriate model for the data (i.e., the types of functions described by the model do not approximate the true underlying function very well). With limited data, it is hard to tell which explanation is correct.