

STAT 830

Statistics versus Probability

I want to begin this course by discussing the difference between Probability Theory and Statistics. Statisticians use the tools of Probability but reason from effects to causes rather than from causes to effects. I want to try to say that again with a bit more detail but still in a vague sort of way.

The standard view of scientific inference starts with a set of theories which make predictions about the outcomes of an experiment as in the following table:

| Theory | Prediction |
|--------|------------|
| A | 1 |
| B | 2 |
| C | 3 |

Now imagine that we actually conduct the experiment and see outcome 2. We **infer** that theory B is correct (or at least that theories A and C are wrong). The question of how much more faith put in B than before is subtle and has been much discussed. As usual theories can easily be falsified – that is, shown to be wrong. But they are only shown to be right in the sense that we try and fail to falsify them. If a theory makes many many correct predictions in many contexts we start to treat it as if it were true; but one wrong prediction demands a rethink.

Now we add **Randomness** to our little table because the outcomes of experiments are not perfectly predictable, even in theory:

| Theory | Prediction |
|--------|-------------------------------|
| A | Usually 1 sometimes 2 never 3 |
| B | Usually 2 sometimes 1 never 3 |
| C | Usually 3 sometimes 1 never 2 |

Now imagine again that we see outcome 2. We now infer that Theory B is probably correct, that Theory A is probably not correct, and that Theory

C is wrong. Notice the precision gained, when Theory C absolutely rules out outcome 2 but outcome 2 actually happens – we can rule out theory C.

That leads me to summarize the difference between Probability and Statistics as follows:

- In **Probability Theory**: we construct the table by computing likely outcomes of experiments. We predict what ought to happen if we do the experiment and some specific theory holds.
- In **Statistics** we follow the inverse process. We use the table to draw inferences from outcome of experiment – deciding how sure we are about which theory is correct. In this course we consider the questions: how should we do draw these inferences and how wrong are our inferences likely to be? Notice: our task is hopeless unless different theories make different predictions – see future discussions of *identifiable* models.