Statistics, Probability, and Inference

I will begin by trying to give working definitions of these three terms.

Statistics concerns the collection, analysis, and presentation of data. Good statistics aims to get the maximum return from the available data, so a good scientist is always interested in it. A common complaint in all areas of science is "we do not have enough data", so we must endeavour to extract as much as possible from that which is available. This is especially true in astronomy where it is often not possible to make repeat observations to get additional data.

The usual situation in science is that we have some (limited) data and we want to use it to learn something about the Universe. For example, we wish to measure the value of some quantity or to test whether a particular model fits the data. This process of forward modelling involves making predictions; starting from a model or a known set of laws and initial conditions calculating the outcome. These predictions are almost always uncertain and probability is the mathematical discipline concerned with quantifying the likelihood that certain events will occur.

We then have to reason backwards in order to draw conclusions about our model; i.e. we are solving the "inverse problem". This is inference. Inference can be a purely logical argument leading to a certain conclusion; however, in the sciences, inference is almost always probabilistic with the observations providing a certain amount of evidence for or against a particular conclusion. It is our task to quantify this evidence.

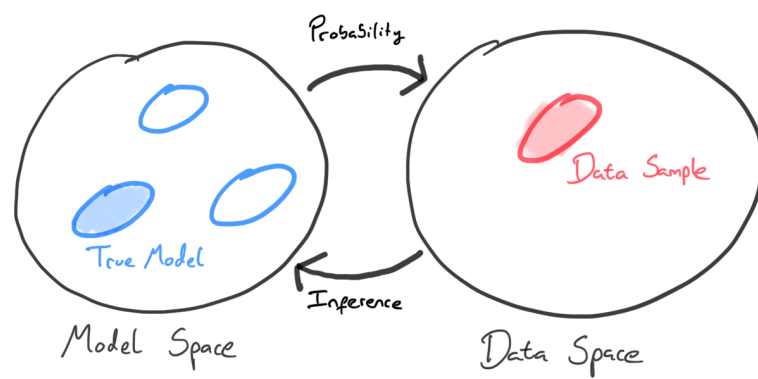The roles of probability and inference are illustrated in the cartoon in Fig. 1

Figure 1: An illustration of the roles of probability and inference.