Introduction & summary

Observations are used to make judgments about reality, how it was, is, or will be under the influence of actions that on their turn can depend on these judgments. That such judgments often differ, shows that, in general, infallibility is impossible. This is not surprising, because induction is a metaphysical affair going beyond ordinary logic. To let the data speak, statisticians use probability models with unknown parameters. It is their task to express the (un)certainty about the unknown of interest, given observations and model. It is obvious that any (mathematical) model will deviate from reality. But also after choosing the model many choices have to be made. These choices will more or less influence the conclusion. This fact is often hidden by referring to only one method, going by the name of a person or a principle. Within a scientific context, however, it may be important to discuss a variety of possibilities such that the ultimate choice can be regarded as well considered.

In this thesis the attention is restricted to inductive reasoning using a mathematical probability model. A statistical procedure prescribes, for every theoretically possible set of data, the inference about the unknown of interest. The ideal is, of course, that such a procedure assigns the inference that coincides with reality, at least for the data at hand. This, however, cannot be guaranteed in the given context. That is why one characterizes procedures by their theoretical performance, which depends on the value of the parameter, under hypothetical repetition. As statements about the true value of the unknown parameter of a probability model are hardly ever directly verifiable or falsifiable, one has to look for other means of evaluation. In this respect it will be investigated whether a procedure displays a systematic error, and whether the inferences it produces are on the average close enough to the actual value of the parameter. The latter can be measured by assigning, to each possible combination of an inference and a theoretically possible value of the parameter, a loss in the form of a score between 0 (perfect) and 1 (totally wrong).

Such a loss function is an instrument to control behavior. It has to be such that minimization of the expected loss leads to reliable inferences. That is, inferences that do not display a systematic error and are close to the true value of the parameter. A loss function with this property is said to be a proper loss function. As there exist many proper loss functions it has to be investigated to what extent the behavior of the statistician depends on the exact form of the loss function. Of course, given a loss function, one tries to develop procedures which, on the average, give rise to small losses; this, preferably independent of the exact form of the loss function. The mathematical expectation of loss, for a
certain value of the parameter, is called the *risk* of a procedure. It is not hard to see that it is impossible to find procedures that have minimum risk for all values of the parameter: local risk minimization goes at the expense of an increase of risk for other values of the parameter. Hence, it is the ultimate goal to find procedures that have an acceptably small risk, for *all* values of the parameter and with respect to *all* proper loss functions.

This thesis consists of two parts. First the problem of assessing the truth or falsity of a statistical hypothesis is considered. Next the attention is shifted towards the more general problem of giving a distributional inference about an unknown real-valued parameter. A central theme is that most of the ideas and concepts that arise from studying the first problem can be translated to the latter.

**The truth or falsity of statistical hypotheses** A statistical hypothesis is a statement about the true value of the unknown parameter. The statement has to be identifiable in the sense that its truth or falsity can be established if the unknown parameter is revealed. A statement about the truth of a hypothesis can be given in the form of a number between 0 and 1, which denotes the ‘chance’ that the hypothesis is true; close to 1 (or 0) expresses that there is great confidence that the hypothesis is true (false), about 1/2 that nothing can be said. Notice that such *epistemic* probability does not correspond to a relative frequency, and hence should not be confused with an ordinary probability. Such a chance, however, is an intuitive way to make the assessment. Historically there are roughly two schools that provide methods to express the (un)certain about the truth or falsity of a hypothesis in the form of a number between 0 and 1: the subjectivistic (Bayesian) school which propagates the use of posterior probabilities, and the objectivistic (frequentist) school which, with some constraints, propagates the use of P-values. To judge whether or not a procedure displays a systematic error, the concepts of *weak* and *strong unbiasedness* are used. It turns out that, in general, Bayesian methods do display a systematic error in this sense, but have a rather small risk. This in contrast with frequentist methods that do not display a systematic error, though they do so at the expense of a slightly larger risk. In the thesis a problem-oriented approach is advocated. For each problem a comparative analysis of various procedures is suggested. In particular the classical problem of testing a $\chi^2$ distribution against the alternative of a noncentral $\chi^2$ distribution is investigated in this way.
Distributional inferences about unknown parameters  After investigating theory to estimate truth or falsity of statistical hypotheses, one is prepared to make distributional inferences about the true value of the unknown real-valued parameter. This means that the opinion about the true value of the parameter, given the observations, is expressed in the form of a probability distribution on the real line. The making of distributional inferences is commonly associated with the Bayesian approach. The unknown parameter itself is then regarded as the outcome of a random variable on the parameter space. The distribution of this random variable is specified such that the joint distribution of the unknown and the observed variables is determined. Distributional inferences can be obtained by applying Bayes's Theorem: the conditional distribution on the parameter space, given the observations, is computed. Distributional inferences can, however, also be made within the framework of a frequentist approach. The family of confidence intervals, of different levels for some observation, can usually be identified with a distributional inference. Consequently, the construction of distributional inferences can be done analogously to the construction of confidence intervals. This gives a direct link to the theory of assessing truth or falsity of statistical hypotheses: the value of distribution function of a distributional inference in a given point on the real line expresses the chance that the true value of the unknown parameter is smaller than this point, i.e., it is an estimate of the truth of the corresponding hypothesis. By using this relation, concepts for assessing truth or falsity of statistical hypotheses – think of, e.g. weak and strong unbiasedness – can be translated into similar concepts for making distributional inferences about real-valued parameters. Similarly, results that hold for methods for constructing epistemic probabilities can be extended to results for making distributional inferences. If such distributional inferences are constructed by using posterior probabilities then results cohere with other ones in the Bayesian approach. The use of P-values leads to distributional inferences that correspond to Fisher’s fiducial argument. This relation between P-values and fiducial distributions gives the possibility to generate fiducial inferences by using results form the Neyman–Pearson theory of testing statistical hypotheses. Besides that, it gives the opportunity to compare fiducial and Bayesian inferences. Again a problem-oriented approach will be adopted. In particular the problem of making distributional inference about the mean of a log-normal distribution will be studied in detail. The fiducial method provides a new procedure that compares favorably with existing ones.