Excerpts from Stanford's *Encyclopedia of Philosophy* article on the Philosophy of Statistics article at http://plato.stanford.edu/entries/statistics/

Now what does the probability function mean? The mathematical notion of probability does not provide an answer. The function P may be interpreted as

physical, namely the frequency or propensity of the occurrence of a state of affairs, often referred to as the chance, or else as

epistemic, namely the degree of belief in the occurrence of the state of affairs, the willingness to act on its assumption, a degree of support or confirmation, or similar.

Physical probability and classical statistics

In the sciences, the idea that probabilities express physical states of affairs, often called chances or stochastic processes, is most prominent. They are relative *frequencies* in series of events or, alternatively, they are tendencies or *propensities* in the systems that realize those events. More precisely, the probability attached to the property of an event type can be understood as the frequency or tendency with which that property manifests in a series of events of that type.

The notion of physical probability is connected to one of the major theories of statistical method, which has come to be called *classical statistics*. It was developed roughly in the first half of the 20th century, mostly by mathematicians and working scientists like Fisher (1925, 1935, 1956), Wald (1939, 1950), Neyman and Pearson (1928, 1933, 1967)… Physical probability cannot meaningfully be attributed to statistical hypotheses, since hypotheses do not have tendencies or frequencies with which they come about: they are categorically true or false, once and for all. Attributing probability to a hypothesis seems to entail that the probability is read epistemically.

Classical statistics is often called *frequentist* … This leads to a central problem for frequentist probability, the so-called *reference class problem*: it is not clear what class to associate with an individual event or item … Since classical statistics employs non-trivial probabilities that attach to the single case in its procedures, a fully frequentists understanding of statistics is arguably in need of a response to the reference class problem.

Epistemic probability and statistical theory

Very roughly speaking, epistemic probabilities can be doxastic, decision-theoretic, or logical.

Probabilities may be taken to represent *doxastic* attitudes in the sense that they specify opinions about data and hypotheses of an idealized rational agent. The probability then expresses the strength or degree of belief, for instance regarding the correctness of the next guess of the tea tasting lady. They may also be taken as *decision-theoretic*, i.e., as part of a more elaborate representation of the agent, which determines her dispositions towards decisions and actions about the data and the hypotheses. Oftentimes a decision-theoretic representation involves doxastic attitudes alongside preferential and perhaps other ones. In that case, the probability may for instance express a willingness to bet on the lady being correct. Finally, the probabilities may be taken as *logical*. More precisely, a probabilistic model may be taken as a logic, i.e., a formal representation that fixes a normative ideal for uncertain reasoning. According to this latter option, probability values over data and hypotheses have a role that is comparable to the role of truth values in deductive logic: they serve to secure a notion of valid inference, without carrying the suggestion that the numerical values refer to anything psychologically salient.

The epistemic view on probability came into development in the 19th and the first half of the 20th century, first by the hand of De Morgan (1847) and Boole (1854), later by Keynes (1921), Ramsey (1926) and de Finetti (1937), and by decision theorists, philosophers and inductive logicians such as Carnap (1950), Savage (1962), Levi (1980), and Jeffrey (1992). Important proponents of these views in statistics were Jeffreys (1961), Edwards (1972), Lindley (1965), Good (1983), Jaynes (2003) as well as very many Bayesian philosophers and statisticians of the last few decades …

Bayesian statistics

The defining characteristic of Bayesian statistics is that it considers probability distributions over statistical hypotheses as well as over data. It embraces the epistemic interpretation of probability whole-heartedly: probabilities over hypotheses are interpreted as degrees of belief, i.e., as expressions of epistemic uncertainty. The philosophy of Bayesian statistics is concerned with determining the appropriate interpretation of these input components, and of the mathematical formalism of probability itself, ultimately with the aim to justify the output. Notice that the general pattern of a Bayesian statistical method is that of *inductivism* in the cumulative sense: under the impact of data we move to more and more informed probabilistic opinions about the hypotheses.

Bayesian inference always starts from a *statistical model*, i.e., a set of statistical hypotheses.

… how do we determine a prior probability? Perhaps we already have an intuitive judgment on the hypotheses in the model, so that we can pin down the prior probability on that basis. Or else we might have additional criteria for choosing our prior.

… the more pressing problem is that different scientists will provide different prior distributions, and that these different priors will lead to different statistical results. In other words, Bayesian statistical inference introduces an inevitable subjective component into scientific method.

It is one thing that the statistical results depend on the initial opinion of the scientist. But it may so happen that the scientist has no opinion whatsoever about the hypotheses. How is she supposed to assign a prior probability to the hypotheses then? The prior will have to express her ignorance concerning the hypotheses. The leading idea in expressing such ignorance is usually the *principle of indifference*: ignorance means that we are indifferent between any pair of hypotheses. For a finite number of hypotheses, indifference means that every hypothesis gets equal probability. For a continuum of hypotheses, indifference means that the probability density function must be uniform.