A contemporary look at the long-lasting Frequentist-Bayesian debate

Prof. Dr. Uwe Saint-Mont

University of Applied Sciences, Nordhausen

Leipzig, 22. Sep. 2011

Interpretation of Probability

1. Frequency / limit of frequencies
2. Personal belief

Standard objections:

1. Frequency is "objective", but does not cover all relevant cases (too narrow)
2. Personal belief is "subjective" and therefore hardly scientific (too broad)

Who cares? Mathematically formalized probability is crucial (Kolmogorov's axioms)

When may I use Bayes' Theorem?

1. Never:
   Subjective prior $P(\theta)$
2. Always:
   Logically sound and automatic way to a probability statement about a latent structure θ given the data x

$$P(\theta|x) = \frac{P(x|\theta)}{P(x)} \cdot P(\theta)$$

3. Sometimes

Isn't this just philosophical quibbling? One must always apply theory in a reasonable manner!

What is random, what is fixed?

1. Random data $P(x)$, but fixed unknown parameter θ
2. Fixed sample x, and random parameter $P(\theta)$
3. Likelihood (Fisher):
   Actual sample x and $L_x(\theta) = P(x|\theta)$, but not $P(\theta)$

Isn't this just a technical detail? Use probability whereever it seems to work well.

A crucial difference

Ingredients of parametric statistics:

- X represents the process generating the data

  $\downarrow$
- x is the actual sample at hand

  $\downarrow$
- θ stands for an latent parameter (or structure)

Sample space inference

Focus on $\underline{X \rightarrow x}$, but not θ

- Consider x relative to all other potential samples
- Emphasis on X and its distribution (given θ)

Interpretation: Prior view / planning before the data are collected.
What may happen?

Parameter space inference

Do not focus on X, but on $\underline{x \longrightarrow \theta}$

- What does the actual x tell about θ?
- X is auxiliary: Its distribution $P_\theta(x)$ only connects the data and the parameter.

Interpretation: Posterior view after the data have been collected.
What did actually happen?

Example: Confidence vs. credibility intervals

"[. . .] a confidence interval is a probability statement about the data, given the parameter, rather than one about the parameter, given the data." (Lindley 2002)

- Confidence Interval: $P_\mu(a \leq \bar{X} \leq b) = 1 - \alpha$.
  Prior perspective, sample space, long run, hypothetical repetition, average (performance) characteristics, procedure, initial precision
- Credibility Interval: $P_{x_1,\ldots,x_n}(a \leq \mu \leq b) = 1 - \alpha$.
  Posterior perspective, parameter space, concrete sample, specific situation, learning from the data, final precision

Example: Confidence intervals

On the one hand, the *procedure* of using the sample mean (or some other measure) to estimate μ could be assessed in terms of *how well we expect it to behave;* that is, in the light of different possible sets of data that might be encountered.

It will have some *average* characteristics that express the precision we *initially* expect, i.e. before we take our data [. . .]

Example: Credibility intervals

The alternative concept of *final precision* aims to express the precision of an inference *in the specific situation we are studying*.

Thus, if we actually take our sample and find $\bar{x} = 29.8$, how are we to answer the question 'how close is 29.8 to μ'?

This is a most pertinent question to ask - some might claim that it is the supreme consideration.

Example: Confidence vs. credibility intervals

Weak interpretation of confidence intervals, in particular:

- "Within the classical approach we must rest on any transferred properties of the long-term behaviour of the *procedure* itself." (Barnett 1999)

Switch of perspective (prior $\longrightarrow$ posterior)

1. Implicit in frequentist statistics: $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$.
2. Explicit in Bayesian statistics (Bayes' Theorem)
3. Fiducial argument (Fisher): Explicit, but without Bayes' Theorem (often debatable)

It should be noted that the aposteriori interpretation of confidence intervals (and thus the implicit fiducial argument and a subconscious switch between aleatory and epistemic probability) was probably centuries old [...] (Hampel 2003)

Focus on X generating x

Elaborated formal prior perspective:

- Experimental Design
- Preferred statistical techniques (e.g. randomization, statistical tests, regression) supersede problem orientation
- Context information is not formalized, and therefore cannot be dealt with in a precise way
- Cult of the single study considered in isolation (Nelder 1999) "Let the data speak for themselves"

$\longrightarrow$ Restricted point of view

Focus on X generating x

Mathematical point of view dominates:

- Emphasis on probability theory, fine properties of random processes, proofs for the most general setting
- Often rigor and formal niceties instead of “useful” mathematics (Hand, Diaconis, von Neumann)

$\rightarrow$ “Mathematistry” (G.E.P. Box): the tendency to produce theory for theory’s sake

Focus on X generating x

Mathematical strategy:

- Weak assumptions $\Rightarrow$ Robust methods
- Weak assumptions $\Rightarrow$ Weak conclusions and interpretations

$\longrightarrow$ No comprehensive theory (toolkit statistician, ad hocery)

Focus on X generating x

Impossible strategy:

- Weak assumptions $\Rightarrow$ Strong conclusions
  only works, if there are further implicit assumptions

- "Many arguments of classical statistics are not fully formalized. A large statistical community implicitly appeal to a variety of conventions rather than presenting rigorous models and assumptions." (Heckman 2005)

  $\rightarrow$ Chasm between statisticians and scientists / mathematical statistics and scientific data analysis

Focus on x and the latent structure θ

Context-oriented strategy:

- Strong assumptions $\Rightarrow$ Strong conclusions
  works if the assumptions are adequate
- Sensitivity analysis, law of decreasing credibility (Manski)
- Explicit Modeling: Full probability modeling, structural equations, causal graphs (Pearl)

$\rightarrow$ Emerging unified formal treatment with a probabilistic core

Focus on x and the latent structure θ

Many major developments in the last decades:

- EDA (Tukey 1977), Leamer (1978, 1983)
- Meta Analysis, Data Mining
- "Extended Bayes": Nonparametric Bayes, Imprecise Probability, Dempster-Shafer theory,...
- Simulation, computational statistics (Gibbs sampling, MCMC)
- Quasi-experiments, control functions, comparability
- Information markets instead of representative samples

$\rightarrow$ Much faster and versatile development in "applied areas" than in orthodox theoretical statistics

Move the data x to center stage

- Li and Vitányi (2008) on traditional statistics: "It is very difficult, if not impossible, to formalize the goodness of fit of an individual model for individual data in the classic probabilistic statistics setting. It is as hard to express the practically important issues in induction in those terms, which is no doubt one of the reasons why contention is rampant in that area."

- Kolmogorov (1974): Reformulation of Statistics independent of probabilistic assumptions
- Nonprobabilistic foundation: finite combinatorics and effective computation
- Focus on the individual data sample and the individual model

Nonprobabilistic Statistics

Unification and sophistication with the help of contemporary information theory:

- $K(x)$ Kolmogorov *complexity* of the data = Length of the smallest computer program generating the data
- Probability of a *single* object x, without reference to a random variable or a distribution
- Model = Compact description of x
- Learning = Modelling = data *compression*
- General decomposition of data:
  Structure & (algorithmic) random component
- General decomposition strategies: MML, MDL

Thank you!

All references can be found in:

Saint-Mont (2011). Statistik im Forschungsprozess.
Eine Philosophie der Statistik als Baustein einer integrativen Wissenschaftstheorie. Physica-Verlag (Springer), Heidelberg