Re: using data for Bayesian prior

Source: http://sci.tech−archive.net/Archive/sci.stat.edu/2004−11/0093.html

---

From: Herman Rubin (*hrubin_at_odds.stat.purdue.edu*)
Date: 11/11/04

Date: 11 Nov 2004 09:59:12 −0500

In article <cmvgag$6ec$1@planja.arnes.si>,
Aleks Jakulin <a_jakulin@@hotmail.com> wrote:
>Geert Verdoolaege wrote:
>> I am not sure why exactly, in Bayesian analysis, it is a bad idea to
>> use information on the measured data for the construction of a prior
>> distribution of the parameters.

For a mechanical Bayesian, this is correct. If you have an
infinitely fast computer with zero cost, you could take any
combination of prior, likelihood, and cost and compute the
Bayes procedure.

>The basic explanation is that you're double−counting the evidence.
>First, you use the data to construct a prior, and then you update the
*>prior with the *same* data.*

There is another way of looking at it which points out that
this is not exactly what is being done. It could be that
the true prior is essentially incomputable, but if one
looks at it as a prior on computable priors, one can
estimate the computable prior from the data, and use this
to compute the action to be taken. In the case of presumably
independent estimation, there are even Bayes empirical Bayes
procedures.

Otherwise, this is not optimal, but it may be close, and
it is possible to get general results here.

>The empirical Bayes procedure uses a part of the data to construct the
>prior, and the other part of the data to update it. This is
>reminiscent of cross−validation, where you use a part of the data to
>construct a model, and the other part to examine how well the model
>fits the data that was not used for the model.

See the above. I do not agree with the above; the entire
data can be used for both, and usually should be.

>However, hard−core Bayesians do not like empirical Bayes very much
>because they see it as an approximation to hierarchical Bayesian
>analysis. There, you don't specify the prior, but instead specify a
>model of the prior's parameters.

What is important is the risk of the procedure. There are those who use computationally simple, but very definitely unreasonable, priors, often based on the data, and look at the problem of approximating the prior. This may or may not be a reasonable way to look at it. When testing a null hypothesis with a reasonable amount of data, the prior probability that the hypothesis is true turns out not to be of importance at all. Looking at the Bayes risk shows this to be the case.

Many of the "hard−core" Bayesians seem to have no qualms in choosing a convenient prior which cannot be reasonable, and often depends on the form of the experiment. This violates the consistency assumptions, which call for minimizing the prior Bayes risk, treating the "prior" as weights, and not as probabilities. The loss and prior cannot be operationally separated, as only the product enters. Using the data to estimate both may well be needed for high−dimensional problems; this includes almost all of the so−called "nonparametric" ones.

−−
This address is for information only. I do not claim that these views
are those of the Statistics Department or of Purdue University.
Herman Rubin, Department of Statistics, Purdue University
hrubin@stat.purdue.edu   Phone: (765)494−6054   FAX: (765)494−0558