Re: Beyond simple penalized regression

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2006−12/msg00728.html

- *From*: hrubin@xxxxxxxxxxxxxxxxxxxxx (Herman Rubin)
- *Date*: 22 Dec 2006 20:39:10 −0500

In article <04ydnb−O−Koi0hbYnZ2dnUVZ_uOmnZ2d@xxxxxxxxxxx>,
Jerry Dallal <gdallal@xxxxxxxxxxxxxxxxxxxx> wrote:

> Herman Rubin wrote:
>
> > In article <nPudnbGNB4ev7xfYnZ2dnUVZ_sC3nZ2d@xxxxxxxxxxx>,
> > Jerry Dallal <gdallal@xxxxxxxxxxxxxxxxxxxx> wrote:
> >
> > > Bob O'Hara wrote:
> > >
> > > > Jerry Dallal wrote:
> > > >
> > > > > JS wrote:
> > > > >
> > > > > > On Dec 20, 1:41 pm, Jerry Dallal <gdal...@xxxxxxxxxxxxxxxxxxxxx> wrote:

....

> > > > > > But is assuming a prior any more dangerous than assuming linearity, or normal residuals, or some other common artificiality.

If done reasonably, it is often no worse than assuming linearity, and probably less bad than making data normal, or using tests of significance.

> > > Yes, and, unfortunately, the question is damning. Linearity, normal residuals, and the like can be checked. "Checking a prior" is an oxymoron. The question is damning because a prior is not something that is "assumed". It *is*.
>
> > No, it is a model assumption. Like any model assumption, one can still check it. One could simply choose another prior, and see if the the results are different if the alternative prior is used.
> >
> > Bob
>
> I'll reply to both Bob and JS here. I'm not sure what kind of Bayesianism you're practicing. I understand how to check model assumptions, but if checking one's prior is not an oxymoron, then the type of Bayesianism you're suggesting becomes nothing more than a self−fulfilling prophecy.

I seem to be the first to derive Bayesian behavior from rationality assumptions only; 59 years ago. My weak axiomatization, published in 1987, shows what is needed for consistency, and that is to consider the overall risk as a positive linear combination of the risks for the various states of nature. This CAN be done, assuming computational feasibility, by computing the posterior distribution for the prior weights and then computing the best action.

But, just as the Gauss−Markov theorem shows that normality
is not of great importance in least squares estimation, it
may be possible to find procedures which are not highly
dependent on the prior, some of them not even formally
Bayesian. Penalized maximum likelihood is typically
equivalent to using a particular unnormalized prior,
so it is really formal Bayes even if claimed otherwise.
In high dimensional cases, and infinite dimensional cases,
such as density estimation, have been considered, one
usually cannot get anywhere without something like this.

....

> > What you fail to realize is that many aspects of a prior do not
> > affect the decision, or not much.

> Oh, I recognize it, but I also recognize that there are many aspects of
> a multidimensional prior that DO matter. Even if it were true that most
> aspects didn't matter, it's the ones that do that are more important. A
> single weak link is all it takes for things to fall apart.

The only practical method of seeing which are important, which
has been carried out and which I strongly recommend, is that
of computing the prior Bayes risk of a procedure. This is often
quite feasible, and shows the problem.

> If priors are always robust or the evidence is always overwhelming, then
> Bayes methods are bogus. Bayes methods are only of value if they can be
> used to update a nontrivial prior to produce a nontrivial posterior.
> I've yet to be convinced that it is possible to assess such priors.

This is more than is needed. Consider the case of spectral
density estimation, in which the current estimation procedures
are close to being formal restricted Bayes estimates with
decreasing priors for the various coefficients. Now one
can treat the case of such formal procedures, using an
empirical Bayes method of estimating the priors, and the
results are good if the assumptions are approximately true,
without making any assumptions on the rate of decrease.

> > For example, if one wishes
> > to test whether a parameter is sufficiently small, if the width
> > of the acceptance interval is small compared to the standard

deviation of the usual estimate, then the testing of a point null will be a good approximation, and here one finds that it is the ratio of the probability of the null to the local density of the alternative, modified by the loss function, which matters.

What the two of you are suggesting seems to be nothing more than a
(subtle?) recasting of noninformative priors.

NO! One cannot use noninformative priors in either of the above situations.

However, JS has already said, "It's more complex but I would kind of agree." I will agree that the more technical mathematical terms one piles on, the harder it is to figure out what's going on! :−)

One does not need to use complicated mathematics. The results are near the surface.

−−

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@xxxxxxxxxxxxxxx Phone: (765)494−6054 FAX: (765)494−0558

.