BAYES IN THE BAÑO: SOME SNARKY REMARKS ON BAYESIAN QUANTILE REGRESSION

ROGER KOENKER

> "We will all be Bayesians in 2020,and then we can be a united profession." D.V. Lindley's 1995 interview with A.F.M Smith, *Statistical Science*.

> "I have lamented that Bayesian statisticians do not stick closely enough to the pattern laid down by Bayes himself: if they would only do as he did and publish posthumously we should all be saved a lot of trouble." [M. Kendall, On the Future of Statistics, JRSS(A), (1968), 131, 182-204].

1. Introduction

To bake a Bayesian π (posterior) I was taught that you needed an $\mathcal{L}$ (likelihood) and a p (prior),[1] so it comes as something of a shock to discover that there are 15,200 web documents[2] employing the phrase "Bayesian quantile regression." Quantile regression would seem to be the very antithesis of a likelihood based procedure, committing the investigator to a parametric model for one paltry conditional quantile function, while professing ignorance, even indifference, about the rest of the *Deus ex machina*.[3]

2. A Most Perplexing Paradox

So what is the attraction? What brings Bayesians to quantile regression like bears to honey? Is it that sweet smell of sin, always so powerful for the priesthood? Or is it that jihadist spirit of the Crusades, intent to recapture Jerusalem from the infidels? Maybe. But more likely it is simply that "Anything you can do, I can do better" confidence immortalized by Ethel Merman in the 1946 Irving Berlin musical *Annie Get Your Gun*.[4] Curiously, this confidence, or should we say creduality, is justified by the familiar Tukeyesque principle that we should, when possible, "borrow strength" from adjacent data. "How much borrowing is prudent?" is always a valid question, however.

If you are interested in the $\tau = 1 - 4/365 \approx 0.99$ quantile of daily maximum ozone concentrations as mandated by the EPA, as in Reich, Fuentes, and Dunson (2009), it is

Version: August 30, 2010. Preliminary, eyes only. Not for circulation, only for squarification. Invited paper for the *Journal of Irascible Results*. This so-called research was partially supported by NSF grant SES-08-50060.

[1] Oh yeah, and probably some data, don't forget the data.

[2] As of April 26, 2010, according to Google.

[3] Data generating mechanism.

[4] If you listen to the song carefully, `http://popup.lala.com/popup/937030210442825432` you will hear that the only thing that both parties to this competitive duet agree upon is that neither one can bake a π.

hard to see how borrowing strength from conditional quantiles far below the median can be very prudent. Unless, of course, you are confident – that word again – that covariate "effects" are invariant across quantiles, and that you know quite a lot about the form of the underlying "error" distribution. An appealling compromise reminiscent of Hill's (1975) early work on inference about Pareto tail exponents, lies in formulating models over relatively narrow bands of quantiles. This approach is illustrated in the reanalysis by Tokdar and Kadane (2010) of recent data on the intensity of tropical cyclones. A challenge with this approach lies in structuring the likelihood and prior so that they do not assert a global influence that would violate the local intent of the model. This challenge is compounded in problems like the cyclone intensity application by the truncation of the observations below a fixed threshold. Borrowing can be good – when the loan can be repaid with interest, but sometimes it requires taking unjustifiable risks. In such cases the spoils to accrue to the circumspect.

3. "Mommy, Where do likelihoods come from?"

Marx famously asserted that he found the Hegelian dialectical method "standing on its head" and proceeded to turn it right side up. Does the Reverend's philosophy of inverse probability need a similar inversion? For Bayesians the likelihood emerges from the mind fully formed, an exemplar of the synthetic *a priori*. Could we place it on a more materialistic footing?

Having once written the cabalistic equation,

$$Q_{Y_i}(\tau|x_i) = g(x_i, \beta(\tau)), \tag{3.1}$$

intending it to describe a single quantile τ, we may, intoxicated by its beauty, be tempted to believe that it holds for the entire world of $\tau \in (0, 1)$. And, having committed this sin of arrogance, we discover that we have also perpetrated a likelihood! The Bayesian response to this realization seems to be: "Zounds! now all we need is a prior for $\beta(\tau)$ and we're in business." The Tukey, exploratory data analysis, reaction is more likely to be: "How do we know g is reasonable, and *especially* how do we know that it is reasonable for all τ? The answer to these agonizing questions leads us into the excrutiating realm of diagnostic testing, as exemplified by Koenker and Xiao (2002). But for those less masochistically inclined, Bayesian introspection about appropriate priors might be a welcome alternative.

Even the most imaginative Bayesian must have days when they need some Virgilian guidance through the circles of such vast *a priori* landscapes. Probability measures on $\{\beta(\tau) \in \mathbb{R}^p | \tau \in (0, 1)\}$, constitute a big space, not quite Hell size, but larger than Texas. Those willing to swing from the empirical branch of the Bayesian family tree might consider the following prescription:

(i) Hold nose,
(ii) Estimate the model (3.1) by frequentist quantile regression[5] on a grid of τ's,
(iii) Compute the usual sandwich estimator of the asymptotic covariance matrix of these estimates on the grid,

[5] Also known as fuqr, or frequently ugly quantile regression, a Sufi term connoting "a state of contentment and renunciation", see `http://chishti.org/sufismquran.htm`

(iv) At each design point, x_i, compute a (smoothed, monotonized if necessary) estimate the conditional quantile function, and the corresponding estimate of the asymptotic covariance function.

(v) Release nose.

At this point you have created, *mirabile dictu*, a Gaussian prior for $\beta(\tau)$. Even more remarkably, however, you have also created a likelihood! We must resist the temptation to call this an empirical likelihood, so we will call it a materialist likelihood. If you insist on a likelihood in the form of a density you will need to differentiate the resulting quantile function with respect to τ, and consider its reciprocal; the covariance function would also need to be modified accordingly via the delta method. But for most "modern" purposes simulation from the prior is what is really needed and this is immediately feasible from the quantile function. Of course there is also the small matter that you have just constructed a prior and a likelihood from the data, you do have some reserved data don't you? If not, take it up with the Reverend next Sunday. A recent manifestation of the materialist likelihood can be found in Reich, Fuentes, and Dunson (2009) where the normal approximation for the fuqr $\hat{\beta}(\tau)$ is used in conjunction with normal priors for coefficients of a Bernstein polynomial expansion to model spatial ozone concentrations. For those wondering whether there are precedents for this sort of apparent heresy in the sacred texts of Bayesianism, one needs to look no further than Jeffreys (1998). Check the index for "median."

References

Hill, B. (1975): "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, 3(5), 1163–1174.

Jeffreys, S. (1998): *Theory of probability*. Oxford University Press.

Koenker, R., and Z. Xiao (2002): "Inference on the quantile regression process," *Econometrica*, 70, 1583–1612.

Reich, B., M. Fuentes, and D. Dunson (2009): "Bayesian Spatial Quantile Regression," `http://www4.stat.ncsu.edu/~reich/SpatialQuantile.pdf`.

Tokdar, S., and J. Kadane (2010): "Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach," preprint.