PAUL'S TOP TEN REASONS TO BE A BAYESIAN

Version of April 2010

1. One can argue that (i) probability is the natural language for describing uncertainty and belief, and (ii) the Bayesian approach is nothing more than the application of probability to whatever problem is at hand, i.e. *what is the probability distribution over what you don't know, conditioned on what you do know.* Thus a Bayesian has *first principles*, and always knows how to proceed with a problem (though the devil may be in the details). In contrast, sometimes being a frequentist is like ordering breakfast in a restaurant: would you like quasi-likelihood, pseudo-likelihood, penalized likelihood, restricted likelihood, profile likelihood, or just regular old likelihood this morning?

2. Using MCMC, we can do inference in virtually any problem, *without relying on asymptotic assumptions.* A lot of frequentist techniques rely on asymptotic arguments to get estimates of precision for point estimates, e.g., the asymptotic normality of ML estimators. These are not universally reliable (think small n, estimates near/on boundary of parameter space, number of variables nearly as big as number of subjects, etc.). Bear in mind that at least in principle the error in MCMC approximations to posterior quantities can be made arbitrarily small by increasing the Monte Carlo sample size (cheap). The error incurred by asymptotic approximations can be made small by increasing the amount of data (expensive).

3. The prior is a flexible tool. If we do have pre-experimental information, then it can be incorporated naturally into a Bayesian analysis. If we don't have any pre-experimental information, then we can use some sort of "flat" prior. The latter approach will usually produce sensible answers, though there are a few problematic issues lurking here. Finally, if we have only vague or poorly-specified prior information, then we can use a robust Bayes approach based on a whole class of priors. I am a particular fan of 'ballpark' priors (which others have more formally termed *weakly informative* priors). In many scientific contexts it is easy to identify implausible magnitudes of parameters *without knowing very much about the particular problem at hand.* For instance, in virtually any biostatistical or epidemiological situation involving binary observational data, odds ratios as big as say 6, or as small as 1/6 aren't very plausible. So a prior could put most (say 95%) of its weight between these values.

4. Bayesian interpretations are clearer. The Bayesian talks about the probability that the unknown parameter lies in a given interval. The frequentist talks about a procedure to generate intervals. If the experiment is repeated over and over again, then a specified percentage of the experiments produce intervals containing the parameter. The Bayesian talks about the probability that a hypothesis is true, in light of the observed data. The frequentist talks about the long-run probability (repeated experiments again) of observing data which are "more extreme" than the data actually observed, given that the hypothesis is true. Clarity here becomes very important. For instance, when comparing a simpler (null) model to an encompassing (alternative) model, one could compute the posterior probability that the null is true, and the classical P-value. Typically the former will be considerably larger than the latter (over a large range of defensible prior distributions).

5. Decision theory dictates that Bayes procedures have good frequentist properties. If we use proper priors then our decisions will be based on *admissible* rules. We also have a long-run calibration: if we repeatedly generate parameters from the prior and data from the model, then in the long-run 95% of the 95% credible intervals will contain the parameter of interest. This calibration property seems under-valued in the literature. Notably, it still holds in scenarios where everything else goes crazy - such as *partially identified* problems.

6. We don't care about data we don't see! In everyday life, we are constantly conditioning on the data we do see. And we revise our previous beliefs as a result. Though we probably aren't using Bayesian updating exactly, it may be a reasonable approximation. In contrast, some frequentist procedures (think P-values in particular) seem overly concerned with data we didn't see.

7. It is impossible for a statistical analysis to be 100% 'objective.' At least Bayesians are honest about this!

8. The Bayesian approach is natural whenever potential data are judged to be partially exchangeable. In particular, there are lots of hierarchical models lurking in the real world (think patients within hospitals within provinces, students within schools within districts, etc.), where *borrowing of strength* is appropriate. These ideas extend to fitting smooth curves and surfaces to noisy data.

9. The Bayesian approach is natural for problems involving unobserved 'latent structure.' Random effects, latent variables, true values of mismeasured covariates, missing data and so on are simply treated as parameters under the posterior distribution. This is especially convenient for MCMC analysis, where these quantities get updated like all the other parameters. This can sometimes result in a Bayesian procedure which is *computationally simpler* than a non-Bayesian analogue.

10. In many situations Bayesian analyses can be regarded at more honest than their classical counterparts in that they take into account more of the uncertainty that is really at play. A frequentist might maximize and subsequently treat the best guess as if it were true in doing susbsequent calculations, which is so-called *plug-in estimation*. In contrast, a Bayesian averages over *a posteriori* plausible values of all the unobservable quantities involved.