On priors and posteriors in statistical inference, Bayesian always and everywhere?

Tore Schweder

November 24, 2016

Statistical inference is about drawing conclusions regarding the state of nature or society based on statistical data, statements equipped with appropriate measures of uncertainties. The Bayesian and the frequentist are the two main schools of statistical inference. They agree that a model specifying the distribution of the data is necessary to make inference regarding the question at hand. They disagree however on whether a prior distribution not based on the data is needed.

As an example assume the data to be yearly macro economic data for USA from 1929 - 1940 on consumption, investment and government spending, and let the question be how much growth in consumption affects investment. Denote by θ this effect size. The model, which the Bayesian and the frequentist agree on, is that the three macro economic time series satisfy a simultaneous linear equations model with normally distributed errors, and only five coefficients in addition to θ. To estimate the parameter in focus, the Bayesian needs a prior distribution for all the six linear parameters in addition to the three error variances. This distribution should ideally be based on knowledge or judgments and should not rest on the observed 1929-1940 data. By combining the apriori distribution and the statistical model a posteriori distribution for θ is obtained by Bayes' formula.

Christopher Sims (Figure 1) used essentially this example to demonstrate the virtues of the Bayesian methodology when he received the *Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel*, for 2011, (Sims, 2012). He chose to use flat and independent prior densities for the six coefficients and for the logarithm of the error variances. The parameter of interest, θ, is assumed positive. The prior distribution is not a probability density since its density integrates to infinity. It is chosen to not color the posterior distribution much. The posterior distribution for θ comes out with a density that nearly is the triangular, $13.3 \cdot (0.15 - \theta)$ for $0 < \theta < 0.15$. The posterior cumulative distribution function is shown in Figure 2.

The prior distribution will color the Bayesian posterior distribution somewhat. There is actually no such thing as a non-informative prior. This lead R.A. Fisher to develop the method of fiducial probability (Fisher 1930), and thus confidence intervals. If 90% of the fiducial probability is within the in-

Figure 1: Christopher Sims.

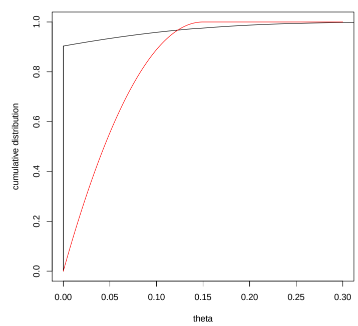

Figure 2: The cumulative confidence distribution for θ (black) together with an approximation to the cumulative posterior distribution obtained by Sims (2012) (red).

terval (a, b), then this is a 90% confidence interval. The modern notion of a *confidence distribution* is that an interval holding a fraction p of the total confidence is a confidence interval of degree $100p\%$ for all $0 \leq p \leq 1$. This amounts to the confidence to the left of the true value of the parameter being uniformly distributed in repeated sampling. In the present, and most other cases, confidence distributions are obtained by simulations, as are Bayesian posteriors. Statistical inference with confidence distributions are discussed in Schweder and Hjort (2016), with many examples including the one discussed here.

I changed Sims' model slightly by assuming $\theta \geq 0$ rather than $\theta > 0$. Otherwise model and data are the same. The resulting cumulative confidence distribution is also displayed in Figure 2. It has a point mass of 0.9003 at zero, and with the rest of the confidence spread out on small positive values. Confidence is really epistemic probability (Schweder, 2016). In view of the data and the model there is about 90% probability that $\theta = 0$ and 96% that it is less than 0.1. This result is obtained without any prior input other than the statistical model, and I claim that the confidence distribution reflects better than the Bayesian posterior our knowledge and its surrounding uncertainty about the effect upon investments of a yearly change in consumption, θ, in USA in the decade before the war.

Sims (2007) claimed that "Econometrics Should Always and Everywhere Be Bayesian". Most of econometrics is however of a frequentist nature — for good reasons since well argued prior distributions seldom are available. The same goes for the other sciences. When the task is to estimate the size of an important parameter, say the climate sensitivity, the method of choice for giving the estimate and its surrounding uncertainty should often be by confidence distribution. Schweder and Hjort (2016) actually claim that exact and optimal confidence distribution in this context is the *gold standard*.

References

[1] Fisher, R.A. (1930) Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26:528-535.

[2] Schweder, T. and Hjort, N.L. *Confidence, Likelihood, Probability; Statistical inference with confidence distributions*. Cambridge University Press. 500 pp.

[3] Schweder, T. (2016) Confidence is epistemic probability for empirical science. To appear in *Journal of statistical Planning and Inference*

[4] Sims, C.A. (2007) Bayesian Methods in Applied Econometrics, or, Why Econometrics Should Always and Everywhere Be Bayesian. Hotelling lecture, June, 2007.

[5] Sims, C. A. (2012). Statistical modeling of monetary policy and its effects [Nobel Prize Lecture in Economics]. *American Economic Review*, 102:1187-1205.