EDITORIAL

This special issue is devoted to some research in Applied Econometrics using exclusively the Bayesian approach to statistical inference. This approach has become more and more popular in view of the computational device known as MCMC (Markov Chain Monte Carlo). Rather than summarizing the essentials of results presented in the issue, we choose to elaborate a bit, for the benefit of a mathematical audience, on two fundamental and useful questions in applying Bayesian statistics.

(i) How to carry out a Bayesian data analysis?

Fortunately, we have a answer handy! Thanks to the prestigeous Text

Bayesian Data Analysis

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin
Chapman and Hall/CRC Press (Third Edition), 2014

which should be our best recommended research (and teaching) material to those who wish to use Bayesian statistics as a tool in uncertainty analysis.

The basic steps to follow in order to conduct a meaningful Bayesian data analysis are spelled out in complete details in this Text. And here they are.

Step 1: Setting up a full probability model (consisting of a sampling model and a prior model). Of course the model should be guided by the problem under investigation.

Step 2: Conditioning on observed data (which consists essentially of deriving the posterior distribution on which inference will be based).

Step 3: Model checking. This is a must step! The results obtained at the end of Step 2 are not “final” yet until a final check is carried out. Since all results depend upon both the suggested sampling model as well as the subjective prior distribution (on the parameter space), these ingredients need to be somewhat validated: For the sampling model, a kind of goodness of fit is required, whereas for the prior distribution, a robust analysis is necessary.

With respect to model checking (see Chaper 6 of the above Text), it is interesting to observe that, while the analysis is completely within the framewaork (and spirit) of Bayesian statistics, the model checking is formulated as a “Fisher- Like” Null Hypothesis Significance Test where the "inference engine" (i.e., the way to jump from data to conclusions) is based on the Fisher's p-value (!) which was avoided in the Bayesian approach to statistics, in the first place, since a p-value only provides a probability of the data given the hypothesis, $P(D|H)$, rather the desired $P(H|D)$ which should be "more logical" to infer about the hypothesis H.

(ii) How to justify the Bayesian approach to statistics?

When we decide to model the epistemic uncertainty of a population parameter by a (subjective) probability measure and then proceed ahead with it, we are in a quite "comfortable" situation since further inference is based on (posterior) probability distributions. However, not all statisticians are Bayesians! since lots of them are not comfortable with their own or others' prior distributions. Thus, the best possible way to "justify" the Bayesian approach is to ask "Is there some way to justify, in some cases, the legitimacy of modeling epistemic uncertainty by probability measures?" For this, we reproduce the famous theorem due to B. De Finetti.

De Finetti Representation Theorem

If $X_n, n \geq 1$, is an exchangeable (infinite) sequence of random variables, then there exist:

(i) a parametric model $f(x|\theta)$, labeled by some parameter $\theta \in \Theta$,

(ii) a probability distribution F on Θ

such that the joint density of $(X_1, X_2, ..., X_n)$ is given as

$$f(x_1, x_2, ..., x_n) = \int_\Theta \prod_{i=1}^n f(x_i|\theta) dF(\theta)$$

meaning that, any finite subset of the exchangeable sequence is a random sample of some model $f(x|\theta)$ and there exists a prior distribution F on Θ which describes the initial information about the parameter which labels the statistical model.

This representation theorem brings out the fact that if observations are exchangeable, then they must indeed be a random sample from some model and there must exist a prior probability distribution for θ so that it is legitimate to consider the parameter as a random variable (with this distribution).

Note that the theorem is only an existence theorem: it does not specify the model and it never specifies the desirable prior distribution!

To be "concrete", we consider the simplest case, with an elementary proof.

Theorem. *Let $X_n, n \geq 1$ be a infinite sequence of* ***binary*** (0-1-*valued*) *exchangeable random variables* (*defined on* $(\Omega, \mathcal{A}, P)$). *Then* ***there exists*** *a distribution function F on $[0,1]$ such that the joint density of $(X_1, X_2, ..., X_n)$ has the form*

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} dF(\theta).$$

Proof. Let

$$f(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$

for $x_i \in \{0,1\}$. Then, by exchangeablity, we have

$$P(X_1 + X_2 + ... + X_n = \sum_{i=1}^n x_i = s_n) = \left(\begin{array}{c} n \\ s_n \end{array} \right) f(x_1, x_2, ..., x_n)$$

where $S_n = X_1 + X_2 + ... + X_n$, and $s_n = x_1 + x_2 + ... + x_n \in \{0, 1, ..., n\}$.

This is so, since, for example, for $n = 3$, and $x_1 = 1, x_2 = 1, x_3 = 0$, we have

$$\begin{aligned}(S_3 = 2) =& (X_1 = 1, X_2 = 1, X_3 = 0) \cup (X_1 = 1, X_2 = 0, X_3 = 1) \\ & \cup (X_1 = 0, X_2 = 1, X_3 = 1)\end{aligned}$$

in which, e.g.,

$$P(X_1 = 1, X_2 = 0, X_3 = 1) = P(X_1 = x_{\sigma(1)}, X_2 = x_{\sigma(2)}, X = x_{\sigma(3)})$$

with $\sigma : \{1,2,3\} \to \{1,3,2\}$, and thus equal to $f(1,1,0)$, recalling that exchangeability means (X_1, X_2, X_3) is equal in distribution to $(X_{\sigma(1)}, X_{\sigma(2)}, X_{\sigma(3)})$, i.e.,

$$\begin{aligned}f(x_1, x_2, x_3) &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= f(X_1 = x_{\sigma(1)}, X_2 = x_{\sigma(2)}, X_3 = x_{\sigma(3)}).\end{aligned}$$

Thus,

$$f(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}) = \left(\begin{array}{c} n \\ s_n \end{array} \right)^{-1} P(S_n = s_n).$$

For $0 \leq s_n \leq n \leq N$ (finite), consider $S_N \in \{0, 1, ..., N\}$, so that

$$P(S_n = s_n) = P((S_n = s_n) \cap \Omega) = P((S_n = s_n) \cap (\cup_{j=0}^N (S_N = j))$$

$$\sum_{j=0}^{N} P(S_n = s_n | S_N = j) P(S_N = j).$$

But $P(S_n = s_n | S_N = j) = 0$ for

(a) $j < s_n$, since $S_N = \sum_{i=1}^{N} x_i$, with $x_i \in \{0,1\}$, so that $S_N \geq S_n$ for $N \geq n$.

(b) $j > N - (n - s_n)$, since $(n - s_n)$ is the number of i such that $x_i = 0$ in $s_n = \sum_{i=1}^{n} x_i$, so that $N - (n - s_n)$ is the maximum number of $1's$ in $S_N = s_n + \sum_{i=n+1}^{N} x_i$.

Thus (taking also exchangeability into account), we have

$$P(S_n = s_n) = \sum_{j=s_n}^{N-(n-s_n)} P(S_n = s_n | S_N = j) P(S_N = j).$$

Now observe that

$$P(S_n = s_n | S_N = j) = \frac{\begin{pmatrix} j \\ s_n \end{pmatrix} \begin{pmatrix} N - j \\ n - s_n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

which is the hypergeometric distribution with parameters (N, j, n). Thus,

$$P(S_n = s_n) = \sum_{j=s_n}^{N-(n-s_n)} \frac{\begin{pmatrix} j \\ s_n \end{pmatrix} \begin{pmatrix} N - j \\ n - s_n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} P(S_N = j)$$

and hence

$$f(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}) = \begin{pmatrix} n \\ s_n \end{pmatrix}^{-1} \sum_{j=s_n}^{N-(n-s_n)} \frac{\begin{pmatrix} j \\ s_n \end{pmatrix} \begin{pmatrix} N - j \\ n - s_n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} P(S_N = j).$$

If we use the notation $(M)_m = \frac{M!}{(M-m)!}$, then

$$f(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}) = \begin{pmatrix} n \\ s_n \end{pmatrix}^{-1} \sum_{j=s_n}^{N-(n-s_n)} \frac{\begin{pmatrix} j \\ s_n \end{pmatrix} \begin{pmatrix} N - j \\ n - s_n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} P(S_N = j)$$

$$= \sum_{j=s_n}^{N-(n-s_n)} \frac{(j)_{s_n}(N-j)_{n-s_n}}{(N)_n} P(S_N = j).$$

Define the distribution function on $[0,1]$ by

$$F_N(.) : [0,1] \to [0,1], F_N(\theta) = \sum_{\{j: \frac{j}{N} \leq \theta\}} P(S_N = j),$$

i.e., $F_N(.)$ is a step function with jump of size $P(S_N = j)$, at $\theta = \frac{j}{N}$, $j = 0, 1, ..., N$.
Then, by writing $j = \theta N$, $N - j = (1-\theta)N$, we have,

$$\sum_{j=s_n}^{N-(n-s_n)} \frac{(j)_{s_n}(N-j)_{n-s_n}}{(N)_n} P(S_N = j) = \int_0^1 \frac{(\theta N)_{s_n}((1-\theta)N)_{n-s_n}}{(N)_n} dF_N(\theta).$$

Since this is true for any N, we should let $N \to \infty$! Now, as $N \to \infty$,

$$\frac{(\theta N)_{s_n}((1-\theta)N)_{n-s_n}}{(N)_n} \to \theta^{s_n}(1-\theta)^{n-s_n} = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

uniformly in θ (using the fact that hypergeometric distribution is approximated by binomial distribution).

Finally, $F_N(.)$, $N \geq 1$, is a sequence of distribution functions, and as such, has a convergent subsequence $F_{N_k}, k = 1, 2, ...$, converging to some distribution function F, i.e.,

$$\lim_{k \to \infty} F_{N_k}(\theta) = F(\theta)$$

by Helly's theorem In summary,

$$f(x_1, x_2, ..., x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} dF(\theta). \qquad \square$$

Remark *The general De Finetti's theorem is this. If $(X_n, n \geq 1)$ is exchangeable under P, then there exists a probability measure Q on the space of all distribution functions $\mathcal{F}$ on $\mathbb{R}$, such that the joint distribution of $(X_1, X_2, ..., X_n)$ has the form*

$$\int_{\mathcal{F}} \prod_{i=1}^{n} F(X_i) dQ(F)$$

Sompong Dhompongsa and Hung T. Nguyen
Guest Editors