A MULTI-PURPOSE TABLE FOR BAYESIAN COMPUTATIONS

by T. Pham-Gia and F. Aucoin,

ABSTRACT:

Newcomers to Bayesian Statistics, already hampered by the new jargon as well as by complex computation formulas, often have also to struggle with a non user-friendly software or have to learn another programming language. The very convenient table presented here will allow them, not only to bring most notions back to the marginal and conditional distributions in basic statistics, but to compute their numerical values as well, hence facilitating the learning of the whole new concept. Other uses of the table are also presented.

KEYWORDS: Distribution, Prior, Posterior, Predictive, Marginal, Conditional, Likelihood, Sampling, Bayes Factor, Regression, Excel, R **.**

AUTHORS:

Thu Pham-Gia, phamgit@umoncton.ca

François Aucoin, frank.aucoin@gmail.com

EDITOR : Professor Mohammad Fraiwan Al-Saleh, malsaleh@sharjah.ac.ae

1. INTRODUCTION

Bayesian paradigm and methods are being introduced in several academic and professional programs, and to a wide variety of students, both graduates and undergraduates, with most of them being non-statisticians. Problems, issues, suggested approaches, recommended solutions, of different types and scopes, can be found in the recent literature (Gajewski and Simon (2008), Allenby and Rossi (2008), Utts and Johnson (2008)), allowing educators to share their experiences. But some serious pedagogical concerns have been rightly raised previously by Moore (1997).

Although topics covered may differ from one course to another, several new notions need to be presented in these introductory Bayesian statistics courses. Some of them are Bayesian Decision-Theoretic oriented, and these notions include: subjective probability, prior distribution, utility theory, loss function, conditional probability, decision process, maximization-averaging, in the traditional decision-theoretic approach. This approach usually requires more statistical maturity from students, but could be quite rewarding if the class is a motivated one. Topics such as Bayesian Estimation Theory, using various types of loss functions, can be dealt with. The posterior risk, the Bayes risk, the sampling size, using some recent research results in the published literature (Pham-Gia and Bekker (2005)) could also be treated, but, preferably, they should be left to a subsequent course, or an add-on reading course. Naturally, first, the meaning of probability, the difference between frequentist and Bayesian considerations of parameters, and other notions, such as proportionality and sufficiency, have to be discussed.

But another approach deals only with the prior and this is the one we will address in this article. This prior can be obtained by elicitation, using subjective probability, or by an analysis of historical data, leaning towards empirical Bayes methods. Two basic ideas introduced there are: 1) the posterior distribution and 2) the predictive distribution, to be fully grasped by students if further comprehension of the Bayesian approach can be expected. Although explained that they are respectively, the conditional, and the marginal distributions, the majority of students still think that they are completely new and unrelated concepts, especially when the conditional and the marginal distributions are to be obtained from the prior and the likelihood function.

The table that we present here will bring most of the new concepts back to these two basic and closely related notions, and a computer program in Excel (also available in R) will then permit the computation of these distributions, with a high degree of accuracy. Bayesian statistics can then be seen as far less complex than it appears. We emphasize the use of Excel here since Microsoft Excel is “the most widely used spreadsheet program” and most students have already mastered it, often since high school. R, on the other hand, should be learned and used if serious study and work on Bayesian Statistics are to be planned. However, a recently published work by two authors, Heiberger and Neuwirth (2009), entitled “R through Excel”, called RExcel for short, perfectly responds to our point of view since it “seamlessly integrates the whole set of R’s statistical and graphical methods into Excel, minimizing for the students the distraction of learning a new programming language”.

By being able to carry out the computations themselves, students will build up their confidence and strengthen their comprehension of the Bayesian paradigm. This positive fact will likely lead to more applications of Bayesian methods once these students will be in a position to use statistics, and apply statistical methods, later in their professional careers. Only the univariate cases (for parameters) and bivariate regression are treated here, for simplicity, but quite advanced concepts can also be treated with this table. For the multivariate case, and more complex issues, specialized works, such as Albert (2007), can be consulted, although our table can be expanded to handle these cases as well, as our numerical example 2 clearly shows. In fact, we plan to develop this table into a full, convenient software dealing with most Bayesian topics.

In section 2 we will present the table gradually, starting with the elementary case of the discrete bivariate random variable in probability, where marginal and conditional distributions are easily computed. The same table will then be used for the discrete Bayesian paradigm, where distributions now become predictive for the sampling variable and posterior for the parameter. Discretization of both the prior and the likelihood function is used in the continuous case, and the table applies again, with integration being performed using cubic splines. This approach to integration will help students overcome a major obstacle that integration represents, and will provide them a useful tool to perform Bayesian computations. Application of the table to the simple example of a proportion with a beta prior is presented here. Section 3 treats in Excel the numerical example already begun in section 2 and shows the simplicity of the table in this case.

Our easy approach should be contrasted with the more "up-to-date" ones using packages such as **Winbugs**, which require much additional knowledge and work from students, beside a big "leap of faith" on the logic and validity of more advanced probability concepts such as Markov chains, Monte Carlo methods, computer simulation that they have not learned. We do not argue against using these probabilistic and computer tools, only against their presence at this early stage.

The two important notions of sufficient statistics and proportionality of distributions are discussed in section 4. In section 5, the table is shown to apply to basic topics in Bayesian statistics, as we see fit for a first course in this domain. The rest of the article is more technical, but could be ignored as first reading. However, the arguments it contains are essential to grasp the potential of our MP Table. Section 6 presents more advanced features of our table together, with another numerical application, concerning the normal-gamma prior in the bivariate normal conjugacy case. Finally, the MP Formula Table gives the list of main Bayesian models with the distributions of their sufficient statistics and an Appendix gives a quick look on how integrals are handled in our table.

2. THE MULTI PURPOSE TABLE (MP-Table)

2.1 We briefly recall here the basic differences between classical (or frequentist) statistics and Bayesian statistics. However, we will not discuss the pro and cons of using the Bayesian approach, nor do we make any detailed comparison between Bayesian and frequentist statistics, on any specific topic, since these discussions are well covered by numerous existing articles, e.g. Albert (1995), and also because, quite often, they call for advanced reasoning and philosophical arguments. This section is hence written for instructors and readers fairly familiar with Bayesian statistics, but who have not taught this subject yet. It also serves as a reference basis for subsequent sections, where we will concentrate on Bayesian computations at the introductory level.

First, in both frequentist and Bayesian approaches, data related to the random phenomenon under study is represented by a random variable **X**, the distribution of which depends on an unknown parameter, denoted by θ. It is of interest to determine the values of the latter since then we can have more, or complete, information about the phenomenon. But, whereas in classical statistics, distribution parameters are considered constants and sampling distributions lead to confidence intervals based on the frequency approach to probability, in Bayesian statistics, parameters are considered as random variables, with their own distributions, called the prior distributions. We hence have Θ, with its prior distribution $g(\theta)$. When data on the variable **X** are gathered, in classical statistics, they are used to give approximate values of the parameter Θ, via some estimation methods such as the maximum likelihood or the moments method, or to make hypothesis test on some specific values. In Bayesian statistics, they would serve to revise the prior, in order to obtain the posterior distribution. Any inference on the parameter will now be based on this distribution since it incorporates prior knowledge on the parameter and the data gathered.

The (1-α)100% credible interval is the interval of the posterior density having (1-α) probability of containing the parameter, contrary to the frequentist confidence interval, which has a frequency interpretation. Traditionally, we write: Posterior $\propto$ Prior $\times$ Likelihood or $h(\theta \mid X) \propto g(\theta) f(X \mid \theta)$, to mean proportionality, where θ is considered as variable, and **X** as a constant . The predictive distribution of the variable **X** itself, named so because it could serve to forecast future values of **X**, relies on the same relation, but **X** is now considered as variable and Θ the mixing parameter. When the prior distribution $g(\theta)$ is used to derive the predictive, the latter is called prior predictive. Using the posterior distribution of Θ, we have, similarly, the posterior predictive distribution.

We can see from the above that data plays a role which is slightly more central in the Bayesian philosophy, and this fact pleads in favor of Bayesian statistics since students will readily recognize that data are the only really tangible entities they have in any statistical analysis. But computations constitute an important component of the Bayesian process too, and this fact has been a major stumbling block for the acceptance of Bayesian methods in applications. Several significant progresses have been made during the last twenty years, using Monte Carlo simulation and the Markov chain approach. We will not discuss these more advanced methods here, but will present the very convenient approach to Bayesian computations offered by our Multi-Purpose Table. Moore (1997) raised two concerns, among others: Absence of a standard content for introductory Bayesian statistics courses and absence of computational software. This article wishes to address these two concerns that seemed still present, in spite of a few textbooks on the market since the publication of his article.

2.2 **<u>Table for a discrete bivariate distribution</u>:** We start with the elementary rectangular table for the discrete random couple (**X,Y**), where the known joint probability mass function $\{p_{ij}, 1 \le i \le m,\ 1 \le j \le n\}$ is given (Fig.1).

$$
\begin{array}{cccc}
 & x_1 & x_n & \text{Y} \\
y_1 & \left[p_{11} \right. & \left. p_{1n} \right] & p_{1.} \\
y_m & \left[p_{m1} \right. & \left. p_{mn} \right] & p_{m.} \\
\text{X} & p_{.1} & p_{.n} & 1
\end{array}
$$

Fig.1: Joint probability mass function $\{p_{ij}, 1 \le i \le m,\ 1 \le j \le n\}$

We know that the marginal distributions of **X** and **Y** are given by $\{p_{.1}, \ldots, p_{.n}\}$ and $\{p_{1.}, \ldots, p_{m.}\}$ respectively, and that the conditional distributions of $(Y \mid x_j)$ is given by $\{p_{ij} / p_{.j}\}$, obtained by dividing the *j*-th column by its total $p_{.j}$ taken from the marginal mass function of **X**.

Similarly, we have the conditional distribution of $(\mathbf{X} \mid y_i)$ given by $\{p_{ij} / p_{i.}\}$. Marginal means and variances, as well as conditional ones, can be easily computed from these mass functions.

2.3 **Bayesian analysis, discrete case:** These straightforward ideas will now be extended to the Bayesian approach, to the discrete case first. $\mathbf{X}$ retains its nature as the r.v. considered in the sampling model, while Y is changed to the discrete parameter Θ considered as a discrete random variable too, with prior mass function $\{p_{1.},..., p_{m.}\}$.

A) Input: Here, the input table is to be filled out by the user. We have labeled columns $C_0, C_1, ..., C_{n+1}$ and Rows $R_0, R_1, ..., R_{m+1}$ in order to identify the entries (Fig. 2).

- Column C_{00}: Numerical values (to be entered) of the parameter Θ, taken within its definition range.
- Column C_0: Prior mass values (to be entered) of Θ at $\{\theta_i\}_{i=1}^{m}$.
- Row R_{00}: Labels $\{P(x_i \mid \theta)\}_{i=1}^{n}$ for different columns. This row will not change.
- Row R_0 : Numerical values (to be entered) of the variable $\mathbf{X}$, $\{x_i\}_{i=1}^{n}$.

Starting from cell (R_1, C_1), the joint densities $\{p_{i\,j}\}$ are computed as $p_{ij} = P(\Theta = \theta_j) P(\mathbf{X} = x_i \mid \theta_j)$, according to the Bayesian paradigm. Upon completion of this input table, it is activated to give Output table 1 and Output table 2.

	C_{00}	C_0	C_1	…	C_n	C_{n+1}
R_{00}			$P(x_1 \mid \theta)$	…	$P(x_n \mid \theta)$	Θ
R_0			x_1	…	x_n	
R_1	θ_1	$p_{1.}$				
…	…	…				
R_m	θ_m	$p_{m.}$				
R_{m+1}	X					$S=\sum_{j=1}^{n} x_j$

Fig.2: Input Table

B) a) Output table 1 gives the marginal distribution of **X**, called prior-predictive distribution $\left\{ p_{.j} = \sum_{i=1}^{m} p_{ij} \right\}_{j=1}^{n}$ on row R_{m+1}, as shown in Table 1 below.

X	x_1	,…	x_n
Pr	$p_{.1}$	,…	$p_{.n}$

Table 1: Prior Predictive distribution of X.

b) Output table 2: (Fig. 3) is obtained by dividing each column of the input table by its total $p_{.i}$, $i = 1,...,n$, given at $\left(R_{m+1}, C_i\right)$, obtaining hence the associated conditional distribution of Θ, when outcome for X is x_i, also called posterior distribution.

$$\begin{array}{c} \\ \theta_1 \\ \\ \theta_m \\ \end{array} \begin{array}{cc} (\Theta \mid x_1) & (\Theta \mid x_1) \\ \left[\begin{array}{cc} \dfrac{p_{11}}{p_{.1}} & \dfrac{p_{1n}}{p_{.n}} \\ & \\ \dfrac{p_{m1}}{p_{.1}} & \dfrac{p_{mn}}{p_{.n}} \end{array} \right] \\ 1 & 1 \end{array}$$

Fig.3: Posterior distribution of Θ (output table 2).

$$\begin{array}{c} \\ X \mid \theta_1 \\ \\ X \mid \theta_m \end{array} \begin{array}{c} \begin{array}{cc} x_1 & x_n \end{array} \\ \left[\begin{array}{cc} \dfrac{p_{11}}{p_{1.}} & \dfrac{p_{1n}}{p_{n.}} \\ & \\ \dfrac{p_{m1}}{p_{1.}} & \dfrac{p_{mn}}{p_{n.}} \end{array} \right] \end{array} \begin{array}{c} 1 \\ \dots \\ \\ 1 \end{array}$$

Fig.4: Conditional distributions of **X** .

Operating row-wise, we have, similarly, the conditional distributions of **X**, for each fixed value of Θ, as given by Fig 4. Although not much used now, these conditional distributions of **X** can be of interest.

C) The posterior-predictive: Suppose that the observed result is x_k. We then put the *k*-th column of Output table 2, which is the posterior distribution of Θ, in column C_0 of a new input table. Then, Output table 1 related to this input table will give us the posterior-predictive distribution mass function, for $\mathbf{X} = x_k$.

2.4 Bayesian analysis, continuous case: The case where either $\mathbf{X}$ or Θ, or both , is continuous can be dealt with by discretizing its/their domain(s) and performing the same operations as before. This discretization process has to be carried out by hand, although it can be quite simple for distributions defined on finite intervals.

- Column C_{00}: Numerical values $\{\theta_1,...,\theta_m\}$ (to be entered) of the parameter Θ that would adequately cover its domain of variation in the case it is infinite, for example, such that the interval $[\theta_1,\theta_m]$ has $1-\alpha$ probability, with α sufficiently small. These values $\{\theta_i\}_{i=1}^m$ are put in rows R_1 to R_m respectively. Here, a guide covering basic distributions with infinite ranges is available to the students.

- Column C_0: Values of the prior density of θ at $\{\theta_i\}_{i=1}^m$, denoted $\{g(\theta_i)\}_{i=1}^m$.

- Row R_{00}: Labels $\{f(x_i\,|\,\theta)\}_{i=1}^n$ for different columns denoting the values at x_i of the likelihood function $f(x\,|\,\theta)$. This row will not change.

- Row R_0 : Numerical values $\{x_i\}_{i=1}^n$ (to be entered) of the variable **X**, that would adequately cover its definition domain.

- Entries within the table, $(R_i,C_j),\ 1\le i\le m,\ \ 1\le i\le n$, will be products of marginal values.

2.5 Example: We will take, as an example, the simple case when X is binomial, $\mathbf{X} \sim Bin(n,\theta)$, with the prior of Θ being $beta(\alpha,\beta)$. We have the table (Fig. 5):

	C_{00}	C_0	C_1	…	C_n	C_{n+1}
R_{00}			$f(x_1\,|\,\theta)$	…	$f(x_n\,|\,\theta)$	Θ
R_0			x_1	…	x_n	
R_1	θ_1	$g(\theta_1)$				
…	…	…				
R_m	θ_m	$g(\theta_m)$				
R_{m+1}	X					S= $\sum_{j=1}^n x_j$

Fig.5: Input table

where, within the matrix determined by the block (R_1, C_1) to (R_m, C_n), the entries are

$p_{ij} = f(x_j \mid \theta_i) g(\theta_i)$, with $f(x_j \mid \theta_i) = C(n, j)(\theta_i)^j (1-\theta_i)^{n-j}$ (1)

and $g(\theta_i) = (\theta_i)^{\alpha-1}(1-\theta_i)^{\beta-1} / B(\alpha, \beta)$ (2)

(ordinate of the density Beta (α, β) at θ_i), $1 \le i \le m,\ 1 \le j \le n+1$, as shown in Fig.6.

$$
\begin{array}{cccccc}
 & x_0 = 0 & \ldots & x_n = n & \Theta \\
\theta_1 & p_{11} & & p_{1n+1} & p_{1.} \\
 & & & & \\
\theta_m & p_{m1} & & p_{mn+1} & p_{m.} \\
X & p_{.1} & & p_{.n+1} & 1
\end{array}
$$

Fig. 6

Integration is now performed along each column $x_j, 1 \le j \le n+1$, for the integral

$$\int_0^1 prior \times likelihood = \int_0^1 g(\theta) f(x_j \mid \theta) d\theta \qquad (3)$$

and the results are put in row R_{m+1}. A spline approach is used, which is accurate to 4 decimals (see Appendix). It gives the value at j of the prior-predictive distribution $\{p_{.j}\}$, which is the beta-binomial, i.e. X ~ Bebin(n, α, β) .

Next, to obtain the value at θ_j of the posterior distribution (for $\mathbf{X} = x_0$ fixed), the output table 2 is displayed, where entries in each column have been divided by the total of that column, given in the bottom marginal row R_{m+1}. A subroutine permits the drawing of a continuous curve between these values, that can be verified as coming from the distribution $beta(\alpha + x_0, \beta + n - x_0)$, as expected. Another separate subroutine gives the common statistics such as mean, variance, interval $\mu \pm 1.96\sigma$, as well as the 95% highest posterior density (hpd) interval of the parameter, according to the algorithm given by Pham-Gia and Turkkan (1993). It requires, however, a finer division of the domain of Θ so that percentiles can be determined with more accuracy. Section 4 gives a numerical example. Pham-Gia(2004) can be consulted on this topic and on other applications of the beta in Bayesian statistics. Once the observed value x_0 is fixed, the posterior distribution can serve as the new prior, to be put in column C_0, to compute the probabilities associated with $\{\theta_1,...,\theta_m\}$ in a new table, which, in turn, will lead to a new posterior distribution for Θ, and a new predictive distribution for X, called posterior-predictive.

3. A NUMERICAL EXAMPLE

In this section we illustrate the use of our MP-table in the following example in Excel.

3.1 Case of the Bernoulli with beta prior

With the case of the binomial model $\mathbf{X} \sim Bin(n, p)$ considered in section 2.3 , the Excel program shows the input table atop, where the values of $\mathbf{X}$ are inputted in the first row, the values of the proportion *p* are entered vertically in the first column, with the corresponding values of the prior $Beta(3,5)$, as given by (2), in the second column. From the third column, we have in each cell the product: $prior \times likelihood$, where the likelihood is given by (1). Integration, as given by (3), is performed columnwise, giving for each value x_j the value of the predictive distribution of **X** at x_j , on row 13. For *n = 20*, we have the following results:

MULTIPURPOSE TABLE FOR BAYESIAN STATISTICS

Val of Par.	Prob of par.							Values of X													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0,1	0,083755	0,186122	0,1964618	0,131	0,0618	0,022	0,0061	0,0014	0,0002	4E-05	4E-06	4E-07	4E-08	3E-09	1E-10	6E-12	2E-13	6E-15	1E-16	1E-18	7E-21
0,2	0,019834	0,09917	0,23552803	0,3533	0,3754	0,3003	0,1877	0,0938	0,0381	0,0127	0,0035	0,0008	0,0001	2E-05	3E-06	3E-07	2E-08	1E-09	5E-11	1E-12	2E-14
0,3	0,00181	0,015518	0,06318075	0,1625	0,2959	0,4058	0,4348	0,3727	0,2596	0,1483	0,0699	0,0272	0,0088	0,0023	0,0005	8E-05	1E-05	1E-06	8E-08	4E-09	8E-11
0,4	7,96E-05	0,001061	0,00672218	0,0269	0,0762	0,1625	0,2709	0,3612	0,3913	0,3478	0,255	0,1546	0,0773	0,0317	0,0106	0,0028	0,0006	9E-05	1E-05	7E-07	2E-08
0,5	1,56E-06	3,13E-05	0,00029728	0,0018	0,0076	0,0243	0,0606	0,1213	0,1971	0,2628	0,2891	0,2628	0,1971	0,1213	0,0606	0,0243	0,0076	0,0018	0,0003	3E-05	2E-06
0,6	1,06E-08	3,19E-07	4,5485E-06	4E-05	0,0003	0,0013	0,0047	0,0141	0,0344	0,0687	0,1134	0,1546	0,1739	0,1605	0,1204	0,0722	0,0339	0,012	0,003	0,0005	4E-05
0,7	1,45E-11	6,78E-10	1,5032E-08	2E-07	2E-06	2E-05	9E-05	0,0004	0,0016	0,005	0,0128	0,0272	0,0477	0,0685	0,0799	0,0745	0,0544	0,0298	0,0116	0,0029	0,0003
0,8	1,13E-15	9,02E-14	3,4274E-12	8E-11	1E-09	2E-08	2E-07	1E-06	9E-06	5E-05	0,0002	0,0008	0,0024	0,0059	0,0117	0,0188	0,0235	0,0221	0,0147	0,0062	0,0012
0,9	8,51E-23	1,53E-20	1,3089E-18	7E-17	3E-15	8E-14	2E-12	3E-11	5E-10	6E-09	5E-08	4E-07	3E-06	2E-05	8E-05	0,0003	0,0008	0,0016	0,0024	0,0023	0,001
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pred dist	0,011861	0,032879	0,05251246	0,0683	0,0812	0,0907	0,0958	0,0961	0,0921	0,0845	0,0744	0,0628	0,0507	0,039	0,0283	0,0192	0,012	0,0067	0,0032	0,0012	0,0003
																					1,0037
									Poster dist for x = 8												
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0,1	7,061177	5,660796	3,74124175	1,9189	0,7618	0,2424	0,0638	0,0141	0,0027	0,0004	6E-05	7E-06	7E-07	7E-08	5E-09	3E-10	2E-11	9E-13	3E-14	1E-15	2E-17
0,2	1,672155	3,016195	4,48518396	5,176	4,6238	3,3097	1,9589	0,9763	0,4141	0,1504	0,047	0,0126	0,0029	0,0006	0,0001	1E-05	2E-06	2E-07	2E-08	1E-09	7E-11
0,3	0,152634	0,471974	1,20315745	2,3802	3,645	4,4728	4,5382	3,8775	2,8191	1,7555	0,9401	0,4338	0,1727	0,0592	0,0175	0,0044	0,0009	0,0002	3E-05	3E-06	3E-07
0,4	0,006711	0,032282	0,12801122	0,3939	0,9384	1,7913	2,8272	3,7575	4,2496	4,1165	3,429	2,4616	1,524	0,8131	0,3731	0,1466	0,049	0,0138	0,0032	0,0006	9E-05
0,5	0,000132	0,000952	0,0056611	0,0261	0,0934	0,2674	0,633	1,2619	2,1407	3,1104	3,8864	4,1849	3,8864	3,1104	2,1407	1,2619	0,633	0,2674	0,0934	0,0261	0,0057
0,6	8,97E-07	9,71E-06	8,6617E-05	0,0006	0,0032	0,0138	0,049	0,1466	0,3731	0,8131	1,524	2,4616	3,429	4,1165	4,2496	3,7575	2,8272	1,7913	0,9384	0,3939	0,128
0,7	1,23E-09	2,06E-08	2,8625E-07	3E-06	3E-05	0,0002	0,0009	0,0044	0,0175	0,0592	0,1727	0,4338	0,9401	1,7555	2,8191	3,8775	4,5382	4,4728	3,645	2,3802	1,2032
0,8	9,51E-14	2,74E-12	6,5268E-11	1E-09	2E-08	2E-07	2E-06	1E-05	0,0001	0,0006	0,0029	0,0126	0,047	0,1504	0,4141	0,9763	1,9589	3,3097	4,6238	5,176	4,4852
0,9	7,17E-21	4,66E-19	2,4926E-17	1E-15	3E-14	9E-13	2E-11	3E-10	5E-09	7E-08	7E-07	7E-06	6E-05	0,0004	0,0027	0,0141	0,0638	0,2424	0,7618	1,9189	3,7412
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

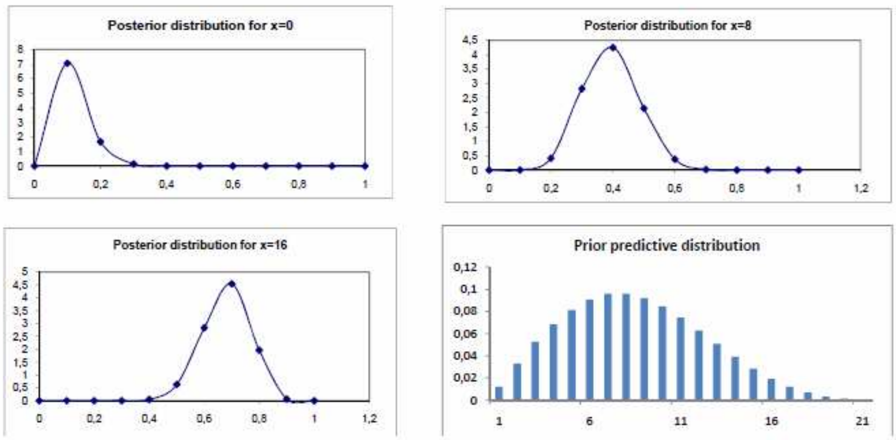

Fig.8 a), b), c), d): Input and output of MP-Table for the binomial case.

a) The prior predictive distribution of **X** is given in row 13, as clearly indicated. Its sum is just slightly above 1, and finer division of the interval for the prior would make it much closer to 1. It is given by Fig.8d, and is indeed $Beta-Binomial(20,3,5)$, as given by the theory.

b) Each posterior distribution, given in the second table, is obtained by dividing the entries in each column of the first table by their sum in row 13. When $x = 8$ we have $p \sim Beta(11,17)$ (Fig. 8 b), as expected.

This very simple example shows the basic method used in our MP-Table: The predictive distribution is (always) obtained *horizontally* in the first table, while the posterior density is (always) obtained *vertically*, under the observed value of the variable, in the second table.

4. SUFFICIENCY AND PROPORTIONALITY

When a sample of several observations is considered it is more convenient to work with sufficient statistics although with our MP-Table, we can use one observation at a time and arrive at the same final result. Hence, the two notions of sufficiency and proportionality need to be introduced and explained since, frequently, non-statistician students find them hard to understand.

4.1 Sufficient statistic and sampling

In the above, we just consider the case of only one observation of the variable X, when studying the posterior distribution of the parameter Θ. When a sample $\{x_1,...,x_k\}$ is taken, the approach is the same since Bayes Theorem allows updating the posterior by considering one observation at a time. Hence the posterior $f(\theta \mid x_1,...,x_k)$ can be obtained in *k* steps. However, some related results cannot be obtained this way, for example, the Bayes Factor (Berger (1999)).

A more convenient approach is to find a sufficient statistic $U = \psi(X_1,...,X_k)$, with computable density, such that the likelihood $f(u \mid \theta)$ has also a computable form. In the beta case above, sampling is really Bernoulli, with $x_i = 0$ or $x_i = 1$, and by taking $U = \sum x_i$, we have the binomial distribution for U. The MP-Table will then be used with Θ and *U* (instead of X). In a special table called the MP-Formula Table, we give expressions of the likelihood, of the sufficient statistics, and of the predictive distribution of *U*. This table is further discussed in section 6.1 and Table 1.

4.2 Normalization and MP-Table results

In all the cases presented above, we have considered exact expressions of the prior densities and also exact expressions of the likelihood functions. The numerical results obtained for the posterior and predictive distributions are then exact. However, very often in Bayesian statistics, we have only the mathematical expression to which the prior density, or the likelihood, is proportional to because the integration constant is unknown in the first case, and the likelihood constant is too complex in the second case. This concept of proportionality is often not very well grasped by students, who feel uncomfortable in its presence. But with the MP-Table we just use all expressions as usual. To obtain the exact conditional density in each column, it suffices to divide each column entry by its total in row R_{m+1}, as before. But the predictive density is obtained by normalization, i.e. by division of all values in Line R_{m+1} by their sum given in cell (R_{m+1}, C_{n+1}). This sum is the product of the two coefficients of proportionality associated with the prior and the likelihood.

5. THE MP-TABLE APPLIED TO TOPICS OF AN INTRODUCTORY COURSE IN BAYESIAN STATISTICS

5.1 Different uses of the MP-Table:

a) The usefulness of the table in the discrete case is hence clear. For the continuous case this depends on the discretization step but its versatility lies in the fact that it can deal with any case, i.e. any sampling model associated with any prior distribution of the parameter. As long as the discretization process gives a good input table, based on adequate truncations of the distributions and a good choice of values $\{x_1,...,x_m\}$ and $\{\theta_1,...,\theta_m\}$, the table is effective. In fact, in an accompanying document available upon request we have provided practical guidelines on how to determine these two sets of values. Hence, numerical values of the posterior and predictive distributions can always be obtained via our table, with the observed values of X used to compute the posterior being either single or multiple, and this should be very helpful to beginners.

b) However, the above table is called the MP-Table because it can also be used in many situations outside Bayesian analysis. For example, when we just take the likelihood function as unity, we have in (C_0, R_{m+1}) the value of the integral $\int g(\theta)d\theta$, with $g(\theta)$ being a density or not. Also, if in (1), **X** and **Y** are regarded as independent parameters, we have in (C_j, R_{m+1}) the numerical value of the continuous mixing of the function $f(x_j,\theta)$ with $g(\theta)$, the latter being the density of the parameter θ. More precisely, let X have density $f(x\,|\,\theta)$, with θ being a random variable with density $g(\theta)$. Then, putting $f(x\,|\,\theta)$ on row R_0 and $g(\theta)$ on column C_0, we will obtain the mixture $\int f(x\,|\,\theta)g(\theta)d\theta$, denoted by $f(x\,|\,\theta)\underset{\theta}{\wedge} g(\theta)$, as the predictive distribution, in row R_{m+1}.

c) The advantages of the table are hence numerous. First, it applies to all cases and hence provides a unifying thread to the whole introductory course. Secondly, it virtually eliminates all computational difficulties, especially those associated with integration. Thirdly, the Bayesian philosophy is constantly put in evidence by emphasizing the prior distribution and observed data, which leads to the posterior distribution, obtained by normalizing the product (prior $\times$ likelihood) via division by the related value of the predictive distribution. Univariate parameters can be conveniently used with the MP-Table, but higher dimensional parameters distributions can also be used, at the expense of more computational complexities

5.2 Topics covered by the MP-Table and by the first course in Bayesian statistics: In Bayesian statistics classroom context, depending on the likelihood function $f(x\,|\,\theta)$, the prior $g(\theta)$ can be non-informative when there is little prior knowledge on the parameter. Usually this will lead to results very similar to those in classical statistics. Non-informative priors, already discussed earlier, will have to be shown to be dependent on the argument used (e.g. Haldane's vs Jeffrey's) but different from the improper priors, with which some statisticians frequently associate with locally non-informative priors. Undoubtedly, this is not an easy task, that requires a lot of thinking from students.

Priors can also be conjugate, leading to closed form formulas for the posterior and predictive, which can be obtained by updating the parameters. Although not necessarily useful or realistic, these cases present much mathematical interest and will permit to verify the numerical values given by the theory with those obtained via our table. These two kinds of priors should be presented in the course.

The four (4) main conjugate cases (a) to d)) are given below, together with the expressions for the prior, and for the likelihood function when dealing with a single observation. They have been verified to work well on the MP-Table. Other topics include: Discrete mixtures, Bayesian tests and Univariate regression (e) to g)). They are often treated as well and are elaborated below, in relation to our MP Table. Since most introductory courses cover the above seven topics only, our table is very suited for such a course. Moreover, it is very convenient for students to remember that the predictive is horizontal in the first table and the posterior is vertical in the second table, under the value x_0 that is observed. The above list contains all topics covered in Bolstad (2004) and, in our opinion, should constitute the essential part of the content of the first course in Bayesian statistics designed for a large audience, not all of them statisticians, and where our MP-Table will well serve its purpose. The opinion we express here represents a consensus on such a content, between the two authors, an instructor and his graduate assistant, and with several other students in the class, hence reconciling several complementary points of view.

a) Normal case: Here, the likelihood is normal (determined by the mean and the precision which is the inverse of the variance), with either the mean, or the precision, unknown.

   i. Mean has normal prior, while precision is a known constant.
   ii. Precision has gamma prior while mean is a known constant.

b) Exponential-Gamma and Poisson-Gamma cases: The likelihood is exponential, or Poisson, whereas the prior of bits parameter is a Gamma.

c) Negative binomial-beta case: The likelihood is negative binomial, the parameter of which has a beta prior.

d) Uniform-Pareto case: The likelihood is uniform on an interval the upper extremity of which has a Pareto prior.

e) Weighted sum of densities as prior (also called discrete mixture): Let us consider the prior density of X which is a weighted average of several densities $f(\theta)=\sum_{i=1}^{n} \alpha_i f_i(\theta)$,

with $\sum_{i=1}^{n} \alpha_i = 1$. We put here in column C_0 the values of f while in row R_0, as usual, we have the likelihood function. It can be shown that the posterior is

$f(\theta \mid X)=\sum_{i=1}^{n} \alpha_i^* f_i(\theta \mid X)$, where $\{\alpha_i^*\}$ are determined by the predictive distribution.

To treat the above topics with our MP-Table, use the prior on the vertical scale and the likelihood on the horizontal scale. The following topics, however, need the use of two MP-Tables.

f) Bayesian statistical tests: the Bayes factor B in favor of the null hypothesis is defined as the ratio of the posterior odds $\frac{p_0}{p_1}$ to the prior odds $\frac{\pi_0}{\pi_1}$, i.e. $B=\frac{p_0}{p_1} / \frac{\pi_0}{\pi_1}$, with $p_0 + p_1 = 1$ and

$\pi_0 + \pi_1 = 1$.

i. Simple hypotheses: Since $B=\frac{f(X \mid \theta_0)}{f(X \mid \theta_1)}$ is the ratio of the two values of the likelihood function at the observed value $X = x_k$ (say), this ratio is immediately available. By using the entries at column x_k and rows θ_0 and θ_1 respectively, we have the ratio of two posteriors

$\frac{\pi_0 f(X \mid \theta_0)}{\pi_1 f(X \mid \theta_1)}$.

ii. Composite hypotheses: Here, the values of θ_0 are in an interval $(a_0,\ b_0)$, while those of θ_1 are in (a_1, b_1), with $(a_0, b_0) \cup (a_1, b_1) = R$. Let $g(\theta)$ be the prior of θ, defined on R. We first normalize the restriction of $g(\theta)/\pi_0$ to $(a_0,\ b_0)$ to make it a density $\rho_0(\theta)$ on this interval. Similarly, we have $\rho_1(\theta)$ a density on (a_1, b_1). We now have $B = \frac{p_0}{p_1} / \frac{\pi_0}{\pi_1} = \frac{\int_{\Theta_0} f(x_0 \mid \theta)\rho_0(\theta)d\theta}{\int_{\Theta_1} f(x_0 \mid \theta)\rho_1(\theta)d\theta}$, where x_0 is the observed value. To compute *B* we use two input tables and proceed as follows:

1. In Input table1, we write in column C_0 the values of $g(\theta)/\pi_0$ in $(a_0,\ b_0)$ and 1 in row R_0. Taking the sum of the column C_1, and dividing all entries by it, we have the density $\rho_0(\theta)$ defined on $(a_0,\ b_0)$. Now, we use input table 2 to deal with the integral in the numerator. It is obtained by putting $f(x \mid \theta)$ in row R_0, as usual, but by putting $\rho_0(\theta)$ just obtained on column C_0, within $(a_0,\ b_0)$. The integral in the numerator is precisely the value under x_k on row R_m in the output table 2.

2. Similarly, we now have $\rho_1(\theta)$ defined on $(a_1,\ b_1)$. We do similarly for $\rho_1(\theta)$ on column C_0, within $(a_1,\ b_1)$, and obtain the value of the integral in the denominator in output table 2 .

3. Their ratio gives us the value of *B*. The posterior probability in favor of H_0 is $p_0 = \frac{\pi_0 B}{\pi_0 B + (1-\pi_0)}$. More details on *B* can be found in Berger (1999).

iii. Case with a prescribed significance level α : Here we need finer values of the posterior distribution, as given by the MP-Table, and the separate subroutines mentioned previously to compute inverse probabilities.

1. Point-null hypothesis: $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$, at α = given degree of significance, based on some given observations. We use Lindley's compromise (Lee (2004, p. 123))by computing the highest posterior density interval already mentioned.

2. One sided null-hypothesis: $H_0 : \theta < \theta_0$, $H_1 : \theta \geq \theta_0$, at α = given degree of significance.

The posterior distribution is given by the MP-Table.

g) Univariate regression:

Independent normal priors can be assigned separately to the slope and the intercept in the bilinear case. Although this is a bivariate prior, we can operate two tables independently of each other.

NOTE: Some of the topics between b) and e) can be dropped if time does not permit to present them all, and Lee (2004) can be consulted for a treatment at a more advanced level.

6. MORE ADVANCED USES OF THE MP TABLE

The following two cases, the continuous bivariate prior and the bivariate regression, require more knowledge and attention from students, but would show them that, complexities aside, the approach is the same.

6.1 **Continuous bivariate priors:** An example of this type of prior is:

α) Normal model with both mean and precision having priors: Here precision has a gamma prior and conditional on the value of this precision, the mean has a normal prior. In the first input table, we put the prior conditional distribution of $\tilde{\mu}$ on the vertical scale while the prior gamma for R is on the horizontal scale. But it needs a third dimension to handle the likelihood. It is to be noted that most software on Bayesian methods use simulation to deal with this problem (Albert (2007, p. 58)), and hence requires some additional knowledge from students. Here, the complete analysis can be done in five (5) steps (from i) to v) below).

i). Prior predictive distributions of $\tilde{\mu}$ and R, separately. Let $R \sim Ga(\alpha, \beta)$ and $(\tilde{\mu} \mid R = r) \sim N(\mu_0, \tau r),\ \tau > 0$. First, the normal-gamma joint prior distribution for the vector $(\tilde{\mu}, R)$ is within the input table itself (and a 3-dimensional graph drawn with these data will help to better understand this prior), whereas the prior predictive of $\tilde{\mu}$ (which is Student) is obtained by integration of the joint density along the horizontal lines. The integral formula derived for this density is:

$$h(\mu) = \int_0^\infty \sqrt{\frac{\tau r}{2\pi}} \exp(-\frac{\tau r}{2}(\mu-\mu_0)^2).\frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1}\exp(-\beta r)dr\,, \tag{4}$$

since in this case the density of μ depends on r. It should also be noted that the distribution for the precision R (which is gamma), is obtained by integration along the vertical direction.

ii). Prior-predictive distribution of X : The third dimension is assigned to the likelihood function, and at each observed value x_0 of X is a table parallel to the first one, which plays the same role as the MP table in the beta case. This table T_{x_0} has at point (μ_i, r_j) the value of the product: $prior \times likelihood$ at (μ_i, r_j, x_0). Hence, integrating w.r.t. μ, and then w.r.t. r gives the value of the prior predictive of X when $X = x_0$, denoted by $\Pr\mathrm{P}(x_0)$. The set $\{\mathrm{PrP}(x_k)\}_{k=1}^{s}, s>1$, gives us the prior- predictive distribution of X , which is a Student distribution. Mathematically, we have the following: Let $X \sim N(\tilde{\mu}, R)$, where $\tilde{\mu}$ and R have densities $h(\mu)$ and $Ga(\alpha,\beta)$ already determined above. Then $f(x)$ is defined point-wise, for $-\infty < x < \infty,$ by:

$$f(x) = \int_{-\infty}^{\infty}\int_{o}^{\infty} f(x\,|\,\mu, r) f(\mu, r) dr d\mu =$$

$$\int_{-\infty}^{\infty}\int_{o}^{\infty} \frac{\sqrt{\tau}}{2\pi}\frac{\beta^\alpha}{\Gamma(\alpha)} r^\alpha \exp\left\{-[\frac{1}{2}\left\{r\tau(\mu-\mu_0)^2 + r(x-\mu)^2\right\} + r\beta]\right\} dr d\mu \tag{5}$$

which can be shown to be $\mathrm{St}\,(\mu_0, \dfrac{\tau}{\tau+1}\dfrac{\alpha}{\beta})$ (6)

with 2α degrees of freedom.

iii). To use our MP-Table, we put one density on each side of the MP-Table, copy this table to another one along the third axis at distance x, $-\infty < x < \infty$, where at the point (μ, r, x_0) the function $f(\mu, r)$ is multiplied by $\frac{\sqrt{r}}{\sqrt{2\pi}}\exp[-\frac{r}{2}(x_0-\mu)^2]$. Performing the integration horizontally and vertically and the value of PrP $(x_0)=f(x_0)$ is at the southeast corner.

iv). Posterior distribution of $(\tilde{\mu},R)$ for $X=x_0$: Dividing the whole table T_{x_0} by $\Pr\mathrm{P}(x_0)$ gives us the Posterior distribution of $(\tilde{\mu},R)$ when $X=x_0$. For a sample of n observations $\{x_i\}_{i=1}^n$, however, the process needs to be applied repeatedly on each x_i to arrive at the posterior distribution $g((\tilde{\mu},R)\,|\,x_1,\ldots,x_n)$.

NOTE: Alternately, if $\mathbf{U}=(U_1,U_2)$ is a sufficient statistic vector when a sample of n observations $\{x_i\}_{i=1}^n$ is considered (see Table 1), with density $f(u_1,u_2)$, we can consider 2 more dimensions instead of only one for X in b), and consider the product $prior*likelihood=\varphi(\mu,R,u_1,u_2)=g(\mu,R)\times f(u_1,u_2)$ in R^4. Integrating w.r.t. μ and r, for a given value of $((u_1)_0,(u_2)_0)$ of (u_1,u_2), we obtain the value of the predictive distribution of (u_1,u_2) at $((u_1)_0,(u_2)_0)$. Dividing $\varphi(\mu,R,u_1,u_2)$ by this value we have the posterior distribution of $(\tilde{\mu},R)$ for $((u_1)_0,(u_2)_0)$.

v). Posterior-predictive distributions: Putting the above posterior distribution in the first table and using it as a prior, by repeating i) and ii) above, we have , respectively, the posterior predictive distributions of $\tilde{\mu}$ and R, when $X=x_0$, and the posterior predictive distribution of X, also when $X=x_0$.

β) Univariate regression: When there is no information on the on the slope and the intercept, non-informative priors are used, and the MP Table can be used as before. An alternate way that can be more convenient is to use closed form results that have been already established with these priors. Table I, the MP Formula Table, provides the necessary details for this second approach.

NOTE on The MP-Formula Table (Table 1): As mentioned in section 4.1, often several independent observations are made in applications instead of just one, and the sufficient statistic, denoted by *U,* will be used for each distribution (see DeGroot (1970)). Our Table I, entitled MP-Formula Table, provides complete descriptions on the nature and distributions of *U* , and also gives technical details for the use of the MP-Table in this case. A version of our table, including detailed instructions, is available upon request, by writing to the first author, indicating either Excel or R. The R version is naturally more powerful and can be used in several other topics of Bayesian statistics not listed above.

6.2 Numerical example 2: As an application of 6.1 α , let us consider the normal likelihood, $X \sim N(\tilde{\mu}, R)$, where both $\tilde{\mu}$ and R (precision) are unknown. We thus take the joint prior density normal-gamma (Case (5c) of MP-Formula Table), here $(\tilde{\mu}, R) \sim N(5,\ 2.4r) \underset{r}{\wedge} Ga(2,5)$, the graph of which is given by Fig. 9a.

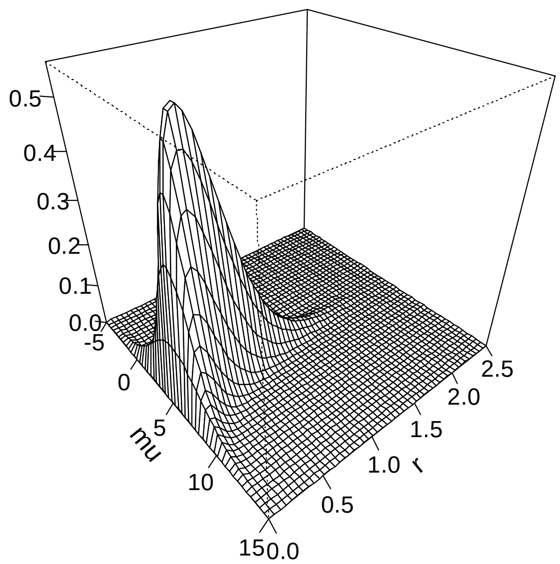

Figure 9a): Joint prior distribution for $(\tilde{\mu}, R)$

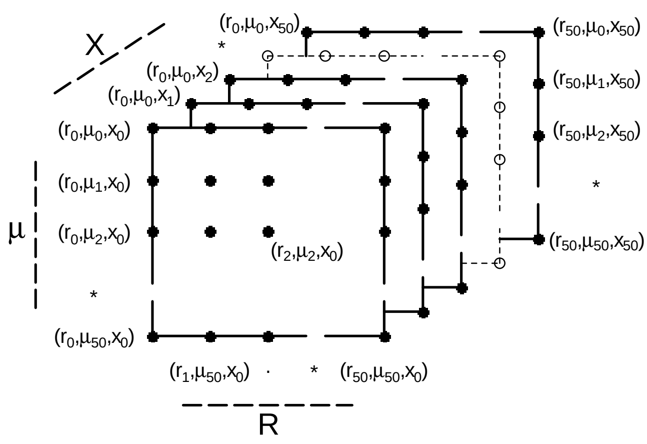

Fig. 9b): Illustration of the 50-by-50-by-50 MP-Table

It is not possible to present the MP table under the same form as in the Excel application because this more complex example necessitates a lot more calculations (multiple MP-Tables, loops, etc.), which would, in turn, require several pages of input. However, handling of the MP-Table in R is also quite simple for less complex applications. We present here some figures and statistics which show that the results obtained by our MP-Table and those from the exact closed form expressions (obtained from the theory) are almost identical. Section 6.1α) is followed closely here, and the results are presented five steps (from 1) to 5) below).

1). First, the MP-Table can be used to determine the prior predictive distributions of $\tilde{\mu}$ (Fig.9d) and R (Fig.9e), which can be shown to be, respectively, $\tilde{\mu}_{prior} \sim St_4(5,.96)$ and $R_{prior} \sim Ga(2,5)$. This is done, with the MP-Table, by first generating a "grid" of the joint prior density $f(\mu,r)$, similar to that used for drawing a perspective plot. Supposing that μ is on the vertical axis, its predictive distribution is obtained by integrating out r across all columns, using the MP-Table. We obtain discrete points $\{\mu_i, f(\mu_i)\}_{i=1}^{m}$. using these values on the MP Table will give the predicted density of $\tilde{\mu}$, as also given by Eq. (4).

2). In this second step we determine:

a. The prior predictive density for X : This time, we will be working in three dimensions. On the vertical axis is μ, on the horizontal axis is r, and on the third axis is x (the likelihood values). The table can now be visualized as having multiple layers of MP-Tables, on each of which the product $prior(\tilde{\mu},R)\times Likelihood\ = f(\mu,r,x_0)$ is computed for the value x_0 of X . So if the previous table for the prior density of $(\tilde{\mu},R)$ was a 50-by-50 table, we now have a 50-by-50-by-50 cube, in which each layer can be viewed as a 50-by-50 table that contains the product $f(\mu,r,x_i)$ (Fig. 9b)) . In order to yield the prior predictive distribution of X at x_i ,we must integrate out μ and r , using the MP Table. More precisely, in the previously described "mesh grid", $f(\mu, r, x_i)$ is now the function on which we need to integrate out both μ and r in order to get $f(x_i)$, value also given by Expression (6). Having obtained all $f(x_i)$'s, for all considered x_i 's , a spline can be fitted to obtain intermediary values, andbthis can be done with our table.

b. Joint posterior distribution of $(\tilde{\mu}, R) | \{x_i\}_{i=1}^{6}$, with $\{x_i\}_{i=1}^{6} = \{4.6, 7.3, 2.4, 5.7, 6.1, 8.3\}$: Here we apply Bayes' theorem sequentially in order to get this distribution. Starting with $x_1 = 4.6$ and having obtained the predictive value of $f(4.6)$, we divide $f(\mu, r, 4.6)$ by $f(4.6)$ in order to obtain the joint posterior $f(\mu, r | x = 4.6)$. Now, this joint posterior distribution is updated with the value $x_2 = 7.3$. We repeat this procedure until all x_i 's are included into the model. Note that the order in which the x_i 's are introduced does not make any difference. Fig. 9c) and 9 d) show that the analytic formulas and our MP-Table give identical results.

3). For the third step we get the posterior predictive distributions for $\tilde{\mu}$ and R by following the same procedure described in a) for the prior predictive, except that this time, the posterior distribution of $(\tilde{\mu}, R)$ obtained above, i.e. $f(\mu, r | \{4.6, 7.3, 2.4, 5.7, 6.1, 8.3\})$ replaces the prior. Fig. 9e) and Fig.9f) illustrate these posterior predictive densities, which are identical to those obtained using closed from formula given by the theory, i.e. $\tilde{\mu}_{posterior} \sim St_{10}(5.52, 2.58)$ and $R_{posterior} \sim Ga(5, 16.25)$.

4). To get the posterior predictive distribution of X , we perform the exact same steps as in 1), substituting the joint prior distribution $(\tilde{\mu}, R)$ by its posterior. Results are presented in Fig. 9g.

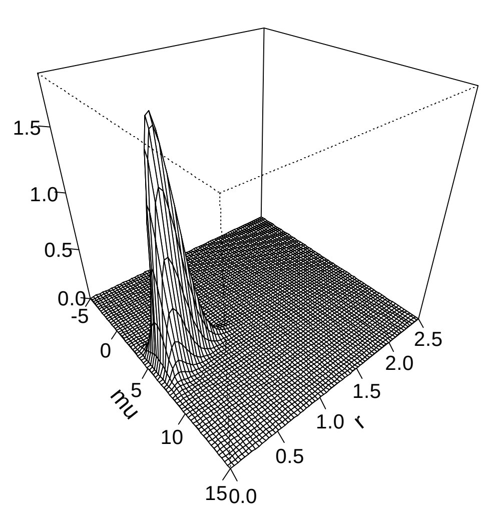

Figure 9c): Joint posterior distribution for $(\tilde{\mu}, R)$, obtained analytically

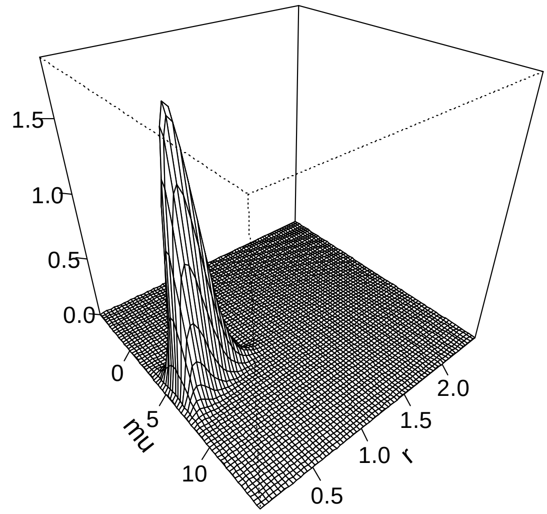

Fig.9d): Joint posterior distribution for $(\tilde{\mu}, R)$, obtained by MP-Table

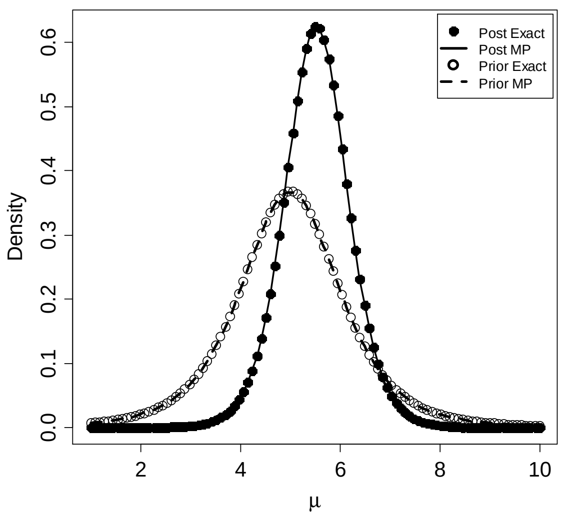

Fig. 9e): Prior and Posterior predictive distributions of $\tilde{\mu}$, obtained with analytical expressions and with our MP-Table

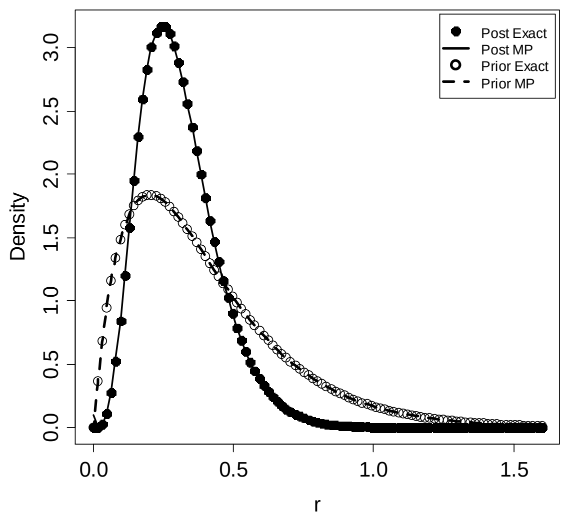

Fig. 9f): Prior and Posterior predictive distributions of R, obtained with analytical expression and with our MP-Table

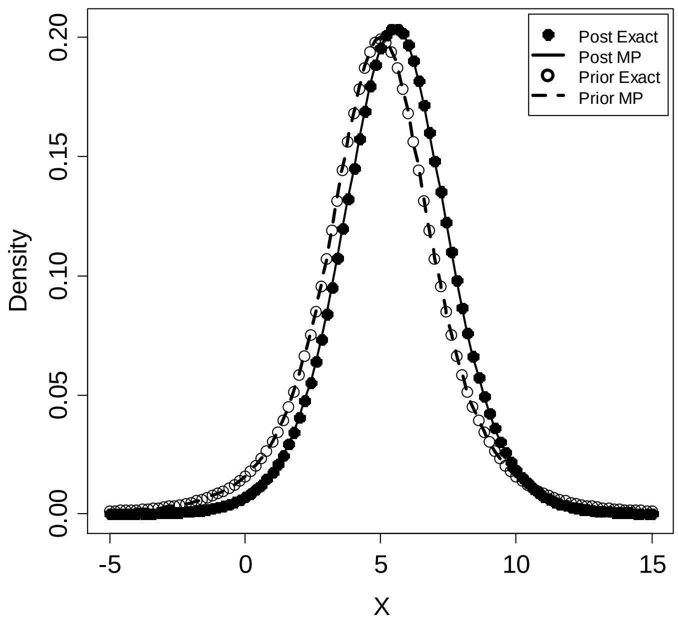

Fig. 9g): Prior and Posterior predictive distributions of X , obtained with analytical expression and with our MP-Table.

5). Now that we have obtained all the prior and posterior distributions of interest with our MP table, it is possible to use them to compute various posterior (prior) statistics and quantities such as the mean, variance, tail probabilities, quantiles and HPD intervals (see Turkkan and Pham-Gia, 1993). This is done by numerical integration and optimization of the posterior spline-fitted distributions, using the MP Table. Again, numerical values for the posterior mean, variance, and quantiles for $\tilde{\mu}$, obtained using either analytical expressions or densities obtained with our MP-Table, are identical, as shown by Table 2. Similarly, Table 3 presents a 90% HDP interval for R that was obtained with both methods.

Table 2 : Exact and MP-Table posterior densities for $\tilde{\mu}$

	Closed form formula	MP Table
Mean	5.5238095	5.523812
Variance	0.4835601	0.483500
1%	3.8048247	3.805099
5%	4.3965111	4.396569
10%	4.6703509	4.670368
50%	5.5238095	5.523737
90%	6.3772681	6.377292
95%	6.6511084	6.651026
99%	7.2427943	7.242479

Table 3 : Exact and MP-Table 90%-HPD intervals for R posterior

	Lower Bound	Mode	Upper Bound
Exact	0.09285445	0.24618138	0.51425366
MP Table	0.0928709	0.2461538	0.5143135

7. CONCLUSION

We have presented a newly designed tool to facilitate computations in Bayesian statistics. The input table given here will allow first the tracking of the prior and sampling model according to Bayesian paradigm. The output tables will help to obtain the predictive and posterior distributions, in either discrete or continuous case. Since these two distributions lay at the heart of the Bayesian paradigm, most topics taught in Bayesian Statistics introductory courses can use this table for computation. Moreover, some applications outside Bayesian statistics can be found for this table as well. To illustrate our approach, two numerical applications are presented: the proportion with a beta prior and the normal case with both mean and precision unknown. Repeated uses, block-wise use, additional dimensions and various extensions of the Multi-Purpose table will permit the computation of results in more advanced topics of Bayesian statistics, and the list of applications of our MP-Table is continuously expanding.

TABLE I

MP-FORMULA TABLE

<u>NOTE</u>: *To use the MP-Table, for the likelihood, use* $f(x;w)$ *if there is only one observation. Otherwise, use* $f(u\,|\,w)$.

N.I. means "non-informative" prior.

1) Model: Poisson, $X \sim Poi(w)$

$$f(x;w) = e^{-x}\frac{w^x}{x!}, \quad x = 0,1,2,..$$

Prior: $W \sim Ga(\alpha,\beta)$, $w > 0$, (N.I. = $\pi(w) = w^{-1/2}$), Sample: $\{x_1,...,x_n\}$

Suff. Stat: $U = \sum x_i$, Likelihood: $f(u\,|\,w) \propto Poi(u\,|\,nw)$.

Post: $W \sim Ga(\alpha+u, \beta+n)$, Pred: $U \sim Pg(\alpha,\beta,n)$

2) Model: Bernoulli, $X \sim Be(p)$

$$f(x;p) = 1 \ or \ 0.$$

Prior: $p \sim beta(\alpha,\beta)$, $0 \le p \le 1$ (N.I.: beta(1/2,1/2)), Sample: $\{x_1,...,x_n\}$

Suff. Stat: $U = \sum x_i$, Likelihood: $f(u\,|\,n,p) = \binom{n}{u} p^u (1-p)^{n-u}, u = 0,1,...,n$

Post: $p \sim beta(\alpha+u, \beta+n-u)$, Pred: $U \sim Bebin(n,\alpha,\beta)$

3) Model: Negative Binomial, $X \sim NB(w,r)$, $r>0$, $x=0,1,2,\ldots$

$$f(x;w,r)=\binom{r+x+1}{r-1}w^{r}(1-w)^{x}.$$

Prior: $W \sim beta(\alpha,\beta)$, $0\le w\le 1$, (N.I.: beta(0 ,1/2)), Sample: $\{x_1,\ldots,x_n\}$

Suff. Stat: $U=\sum x_i$, Likelihood: $f(u\mid r,w)\propto \binom{r+u-1}{r-1}w^{nr}(1-w)^{u}$, $u=1,2,\ldots$

Post: $W \sim beta(\alpha+nr,\beta+u)$, Pred: $U \sim Nbb(\alpha,\beta,nr)$

4) Model: Continuous uniform, $X \sim Un(0,w)$, $0\le x\le w$,

$$f(x)=1/w$$

Prior: $W \sim Pa(\alpha,\beta)$, (N.I.: $\pi(w)=w^{-1}$), Sample: $\{x_1,\ldots,x_n\}$

Suff. Stat: $U=\max\{x_1,\ldots,x_n\}$, Likelihood: $f(u\mid w)\propto Ip(n,1/w)$ for $u<w$,

= 0, otherwise.

Post: $W \sim Pa(\alpha+n,\max\{\beta,x_1,\ldots,x_n\})$, Pred: $U \sim \frac{n}{\alpha+n}Pa(\alpha,\beta)$ if $u>\beta$,

$$U \sim \frac{\alpha}{\alpha+n}Ip(n,\beta^{-1}),\ u\le\beta$$

5) Model: Exponential, $X \sim Exp(w)$, $0<x$

$$f(x,w)=w\exp(-wx)$$

Prior: $W \sim Ga(\alpha, \beta)$, (N.I. : $\pi(w) = w^{-1}$), Sample: $\{x_1, ..., x_n\}$

Suff. Stat: $U = \sum x_i$, Likelihood: $f(u \mid w) \propto u^{n-1} \exp(-wu)$

Post: $W \sim Ga(\alpha + n, \beta + u)$, Pred. $U \sim Gg(\alpha, \beta, n)$

6) Model: Normal, $X \sim N(\mu, r = 1/\sigma^2)$

$$f(x; \mu, r) = \sqrt{\frac{r}{2\pi}} \exp(-\frac{r(x-\mu)^2}{2}), -\infty < x < \infty .$$

a) $R = r_0$, Prior: $\mu \sim N(\mu_0, \tau_0)$ (N.I. : $\pi(\mu) = const.$), Sample: $\{x_1, ..., x_n\}$

Suff. Stat: $U = \sum x_i / n = \overline{x}$, Likelihood: $f(u \mid r_0, \mu) \propto \exp(-\frac{nr_0}{2}(u-\mu)^2)$

Post. $\mu \sim N(\frac{\tau_0 \mu_0 + nr_0 \overline{x}}{\tau_0 + nr_0}, \tau_0 + nr_0)$, Pred. $U \sim N(\mu_0, \frac{nr_0 \tau_0}{\tau_0 + nr_0})$

b) Prior: $\mu_0, R \sim Ga(\alpha, \beta)$, (N.I. : $\pi(R) = r^{-1}$) , Sample: $\{x_1, ..., x_n\}$

Suff. Stat: $U = \sum (x_i - \mu_0)^2$, Likelihood: $f(u \mid \mu_0, r) \propto Ga(n/2, r/2)$

Post: $R \sim Ga(\alpha + \frac{n}{2}, \beta + \frac{\sum (x_i - \mu_0)^2}{2}))$, Pred. $U \sim Gg(\alpha, 2\beta, n/2)$

c) Prior: $R = Ga(\alpha, \beta), (\mu \mid R = r) \sim N(\mu_0, \tau_0 r)$, $\tau_0 > 0$, (N.I. : $\pi(\mu, R) = r^{-1}$) ,

Sample: $\{x_1, ..., x_n\}$

Post. $R \sim Ga(\alpha+\frac{n}{2}, \beta+\frac{\sum(x_i-\bar{x})^2}{2}+\frac{\tau_0 n(\bar{x}-\mu_0)^2}{2(n+\tau_0)}$, $(\mu \mid R=r) \sim N(\frac{\tau_0\mu_0+n\bar{x}}{\tau_0+n}, (\tau_0+n)r)$,

Pred. $\mu \sim t_{2\alpha}(\mu, \frac{\alpha\tau_0}{\beta})$, $R \sim Ga(\alpha,\beta)$, $X \sim St_{2\alpha}(\mu_0, \frac{\tau_0}{\tau_0+1}\frac{\alpha}{\beta})$

7) Regression : A) NORMAL PRIORS : Model $\mu_{Y|x} = \alpha_{\bar{x}} + \beta(x-\bar{x})$,

$$SS_X = \sum(x_i-\bar{x})^2, SS_Y = \sum(y_i-\bar{y})^2,\ SS_{XY} = \sum(x_i-\bar{x})(y_i-\bar{y}),\ B = \frac{\overline{XY}-\overline{X}\,\overline{Y}}{\overline{X^2}-\overline{X}^2},\ A_{\alpha_{\bar{x}}} = \overline{Y}$$

1) Likelihood: 1) for $\alpha_{\bar{x}} \propto \exp(-\frac{(\alpha_{\bar{x}}-A_{\bar{x}})^2}{\sigma^2/n})$, 2) for $\beta \propto \exp(-\frac{(\beta-B)^2}{2\sigma^2/SS_X})$

Prior: a) $\alpha_{\bar{x}} \sim N(m_{\alpha_{\bar{x}}}, s_{\alpha_{\bar{x}}}{}^2)$, Post. $\alpha_{\bar{x}} \sim N(\frac{(1/s^2_{\alpha_{\bar{x}}})}{G}m_{\alpha_{\bar{x}}} + \frac{n/\sigma^2}{G}A_{\bar{x}},\ G)$,

$$G = \frac{1}{s^2_{\alpha_{\bar{x}}}} + \frac{n}{\sigma^2}.$$

b) $\beta \sim N(m_\beta, s_\beta{}^2)$ $(N.I. = flat)$, Post. $\beta \sim N(\frac{(1/s^2_\beta)}{H}m_\beta + \frac{SS_X/\sigma^2}{H}B,\ H)$,

$H = \frac{1}{s^2_\beta} + \frac{SS_X}{\sigma^2}$, with *G* and *H* being precisions.

2) Prediction interval: We first have the posterior distribution of $\mu_{n+1} = \alpha_{\bar{x}} + \beta(x_{n+1}-\bar{x})$ as

$N(m^*_\alpha + m^*_\beta(x_{n+1}-\bar{x}),\ (s^*_\alpha)^2 + (s^*_\beta)^2(x_{n+1}-\bar{x})^2)$, denoted as $\mu_{n+1} \sim N(m^*_{n+1}, (s^*_{n+1})^2)$.

Hence, $\mu_{n+1} \sim N(m^*_{n+1}, (s^*_{n+1})^2)$

$$f(y_{n+1}) \propto \exp(-\frac{(y_{n+1}-\mu_{n+1})^2}{2\sigma^2}).$$

By applying these results to the MP Table, we will have the predictive distribution of Y_{n+1}.

In closed form, we have $(Y_{n+1} \mid x_{n+1}, data) \sim N(m_{\mu}^{*},\ \sigma^2 + (s_{n+1}^{*})^2)$, i.e. the same distribution as the posterior distribution of μ_{n+1}, but with the variance augmented by σ^2.

B) NON INFORMATIVE PRIORS: model $\mu_{Y|x} = \alpha_x + \beta x$, usually we take non-informative priors, $p(\alpha_x) \propto const,\ p(\beta) \propto const,\ p(\sigma) \propto \frac{1}{\sigma}$, the prior is then $p(\alpha_x, \beta, \sigma) \propto 1/\sigma$ and is a trivariate density which depends only on the last variable while the likelihood is $\propto \frac{1}{\sigma^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum \left(y_i - \alpha_x - \beta x_i \right)^2 \right\}$. We can show that the posterior distribution of (α_x, β) is:

$$h\ (\alpha_x, \beta \mid X, Y) \ \propto\ \left\{ (n-2)s^2 + n(\alpha_x - \hat{\alpha}_x)^2 + \left(\beta - \hat{\beta}\right)^2 \sum x_i^2 + 2(\alpha_x - \hat{\alpha}_x)\left(\beta - \hat{\beta}\right) \sum x_i \right\}^{-n/2}$$

, which is a bivariate *t*-distribution. Hence the posterior distributions of α_x and β are univariate *t*-distributions with *(n-2) d.f.* More precisely, we can use the following close form formulas: 1.

$$C_1\{(\alpha_x - \hat{\alpha}_x) \mid (X,Y)\} \sim t_{n-2},\ \text{with}\ C_1 = \sqrt{n \sum (x_i - \bar{x})^2 / s^2 \sum x_i^2}$$

and $C_2\left\{(\beta - \hat{\beta}) \mid (X,Y)\right\} \sim t_{n-2}$, with $C_2 = \sqrt{\sum (x_i - \bar{x})^2 / s}$.

2. Also, the posterior density of the standard deviation σ:

$$h(\sigma \mid X, Y) \propto \frac{1}{\sigma^{n-1}} \exp\left(-\frac{(n-2)s^2}{2\sigma^2}\right)$$ is an inverse gamma.

3. For prediction, we can show that if $y^* = \alpha_x + \beta x^* + u$, then $C_3\{(y^* - \alpha_x - \beta x^*) \mid X, Y\}) \sim t_{n-2}$,

with $C_3 = 1/\{\hat{\sigma}\sqrt{1 + (1/n) + [(x^* - \bar{x})^2 / \sum (x_i - \bar{x})^2]}\}$, where $\hat{\alpha}_x$, $\hat{\beta}$ and $\hat{\sigma}$ are the maximum likelihood estimates of α_x, β and σ respectively.

APPENDIX

INTEGRATION

One of the most confusing and difficult notions for non-math students is integration, especially when it is combined with other mathematical operations. Indeed, the computation of exact values for integral expressions, either in closed form, numerically or by simulation, remains one of the main obstacles in learning Bayesian statistics. In this article, we have taken a simple approach that would allow students to confidently use integration and obtain the (approximate but quite precise) numerical value of practically any integral, without having to consider the complex mathematics associated with it.

A) Simple integral: Suppose we have to compute $\int_{-\infty}^{\infty} f(x)dx.$, with $f(x) \geq 0.$

1) First, we determine the interval (a, b) where lies most of the values of $f(x) \geq 0.$ This means that outside (a, b), $f(x) \leq \varepsilon,\ \varepsilon$ sufficiently small. For example, for $X \sim N(\mu, \sigma^2)$, we can take (a,b) = $\mu \pm 3\sigma$. But for each distribution careful consideration must be made to obtain (a,b).

2) The range (a,b) is now divided into consecutive intervals $\bigcup_{i=o}^{n}(x_i, x_{i+1}),$ with $x_0 = a,\ \ x_n = b$, with n large.

3) We discretize *f*, i.e. compute the set of $\{f(x_i)\}_{i=0}^{n}$ and draw the bar-diagram $\{x_i, f(x_i)\}_{i=0}^{n}$.

4) Between two successive bars $\{x_i, f(x_i)\}$ and $\{x_{i+1}, f(x_{i+1})\}$ we use a cubic spline (which is just a curve of the third degree to smoothly connect two consecutive points) to approximate the values of $f(x)$ in the interval (x_i, x_{i+1}), obtaining a continuous curve f^* on (a,b) which approximates the graph of f. The whole process is obtained by solving a set of equations that give the values of the coefficients of the cubic spline.

5) Integration is now carried out as sums of areas of trapezes based on f^*. If $\{(x_i, x_{i+1})\}_{i=0}^{n-1}$ is used then $Int = \sum_{i=0}^{n-1}(x_{i+1} - x_i) f^*(x_i)$. A finer division of (a,b) would still use f^* (Thisted (1988, 299-300)).

<u>NOTE</u>: Sometimes we will set $x_{i+1} - x_i = \alpha$, and the integral will have a very simple expression as $Int = \alpha \sum_{i=0}^{n-1} f^*(x_i)$, with $n = (b-a)/\alpha$. For a large value of *n*, the value of *Int* obtained is very close to the exact value of the integral.

B<u>) Iterated integral</u>: $\int_a^b \int_c^d f(x,y)dydx$ can be carried out by just using part A) sequentially, first, integrating w.r.t. to y, to obtain a set of points $\{x_i, f^*(x_i)\}$,where $f^*(x_i) = \int_c^d f(x_i, y)dy$, and then applying A) again to that set.

<u>C) For a bivariate function</u> $f(x,y)$, similarly, we start with a rectangle A =(a,b)x (c,d) such that $f(x,y) \le \varepsilon$ outside A.

1) A mesh $\bigcup_{j=1}^{m}\bigcup_{i=1}^{n}(x_i, x_{i+1})\times(y_j, y_{j+1})$ = A is now obtained. The function f is defined at each mesh point $\left(x_i, y_j\right)$.

2) Within $(x_i, x_{i+1})\times(y_j, y_{j+1})$, the product of the two cubic splines that approximate respectively $f(x, y_j)$ and $f(x_i, y)$ gives an approximation of $f(x, y)$.

3) Integration of $f(x, y)$ is carried out like in the univariate case. Here it is a sum of volumes, based either on $(x_i, x_{i+1})\times(y_j, y_{j+1})$, or on a finer subdivision.

ACKNOWLEDGEMENTS

This research was partially supported by NESRC grant A 9249 (Canada). The authors wish to thank the class MATH 6033: Bayesian Statistics, at the Université de Moncton, for the feedback it provided when they tried to implement this approach in class, and to Professor N. Turkkan for his expert advice.

REFERENCES

Albert, J.(2007), "Bayesian Computation with R", 2nd Ed. , New York, Springer.

Allenby, G.M. and Rossi, P.E.(2008), "Teaching Bayesian Statistics to Marketing and Business Students", The American Statistician, 62, 3, 195-198.

Berger, J.(1999), " Bayes Factor", in Update 3, Encyclopedia of Statistics, S. Kotz, Read, C.B. and Banks, D.L., Ed., 20-29, Wiley, N.Y.

Bolstad, W. M.(2004), "Introduction to Bayesian Statistics", New York: Wiley.

DeGroot, M. (1970), "Optimal Statistical Decisions", NY: Mc Graw-Hill, Gajewski, B.J. and Simon, S.D.(2008), "A One-hour Training Seminar on Bayesian Statistics for Nursing Graduate Students", The American Statistician, , 62, 3, 191-194.

Heiberger, R.M. and Neuwirth, E. (2009), R through Excel,NY: Springer.

Lee, P. M.(2004), "Bayesian Statistics, An Introduction", 3rd Ed., London: Hodder Arnold.

Moore, D.S.(1997), "Bayes for Beginners? Some Pedagogical Questions", Advances in Statistical Decision Theory, S. Panchapakesan and. Balakrishnan, N. ,Eds., Birkhauser, 3-17.

Pham-Gia, T. and Bekker, A. (2005), " Sample sizes determination using Bayesian decision criteria under absolute value loss". American Journal of Mathematics and Management Sciences, 25(3,4), 259-291.

Pham-Gia, T. (2004), "Bayesian Inference, in Handbook of the Beta Distribution", A. Gupta and S. Nadarajah, Ed., Boca Raton: CRC Press.

Pham-Gia, T. and Turkkan, N.(1993), "Computation of the Highest Posterior Density Interval in Bayesian Analysis", J. Statist. Comput. Simul., 243-250.

R Development Core Team, R (2009): "A Language and Environment for Statistical Computing, R" Foundation for Statistical Computing, Vienna, Austria, 2009, {ISBN} 3-900051-07-0, http://www.R-project.org.

Thisted, R.A.(1988), "Elements of Statistical Computing", London: Chapman and Hall/CRC.

Utts, J. and Johnson, W.(2008), "The Evolution of Teaching Bayesian Statistics to Nonstatisticians: A Partisan View from the Trenches", The American Statistician, 62, 3, 199-201.

AUTHOR CONTACT INFORMATION

First author: T. Pham-Gia

Department of Mathematics and Statistics, Universite de Moncton, Moncton, New Brunswick, Canada E1A 3E9.

Email address: Phamgit@umoncton.ca

Phone: (506) 8584253; Fax: (506) 8584396

Co-author: François Aucoin, Department of Mathematics and Statistics, Universite de Moncton, Moncton, New Brunswick, Canada E1A 3E9