UNIVERSITAT POLITÈCNICA DE VALÈNCIA
IN-RED 2019
V Congreso de Innovación Educativa y Docencia en Red
Congreso In-Red 2019
UPV, 11 y 12 de julio de 2019
Doi: http://dx.doi.org/10.4995/INRED2019.2019.10369

A Labview® program for illustrating the basic concepts of Bayesian inference*

Juan José Giner-Sanz[1,2], Montserrat García-Gabaldón[2], Emma María Ortega[2], Yang Shao-Horn[1] and Valentín Pérez-Herranz[2]

[1]The Electrochemical Energy Lab, Massachusetts Institute of Technology
[2]Depto. Ingeniería Química y Nuclear, Universitat Politècnica de València

Abstract

Despite the importance of Bayesian inference and the growth of Bayesian research, today, most undergraduate teaching is still based on frequentist statistics. A way of facilitating the introduction of students to the Bayesian world is to strongly reinforce the basic concepts behind the Bayesian philosophy. In this work, a simple Labview® program for reinforcing and illustrating the basic concepts underlying Bayesian inference is presented. This program may be used in a computer lab session, or as an online applet for the students to revise the concepts after the class or in a Massive Open Online Course (MOOC) course.

Keywords: *Bayesian inference, Labview®, Pseudo-random number generation, Biased coin simulation, Bias estimation.*

Resumen

A pesar de la importancia de la inferencia Bayesiana y el crecimiento de la investigación Bayesiana, hoy por hoy, la mayoría de los planes de estudio de grado todavía se basan en la estadística frecuentista. Una forma de facilitar la introducción de los estudiantes al mundo Bayesiano es reforzar los conceptos básicos de la filosofía Bayesiana. En este trabajo, se presenta un programa implementado en Labview® para reforzar e ilustrar los conceptos básicos que subyacen a la inferencia Bayesiana. Este programa se puede usar en prácticas informáticas, o como un applet en línea para que los estudiantes revisen los conceptos después de clase o en un curso online masivo y abierto (MOOC, por sus siglas en inglés).

Keywords: *Inferencia Bayesiana, Labview®, Generación de números pseudoaleatorios, Simulación de una moneda sesgada, Estimación del sesgo.*

*The authors are very grateful to the Generalitat Valenciana and to the European Social Fund, for their economic support in the form of Vali+d postdoctoral grant (APOSTD-2018-001).

1 Introduction

The result known nowadays as the Bayes' theorem can be tracked down to the paper of a nonconformist English minister,the Reverend Thomas Bayes, "An essay towards solving a problem in the doctrine of chances" (Bayes 1763) posthumously published in 1763. This paper contains what is arguably the first detailed description of the elementary probability theory theorem associated with his name today. However, it was the XVIIIth century french scientist, Pierre-Simon Laplace, who introduced a general version of this result in his 1774 paper "Mémoire sur la probabilité des causes par les événements" (De Laplace 1774). This early Bayesian inference, called at the time "inverse probability", was largely used until the 1920s to approach a large variety of problems ranging from celestial mechanics, medical statistics and reliability to jurisprudence (Stigler 1986). Even if the Bayes' theorem has around 250 years of history, and the method of inverse probability that flowed from it dominated statistical thinking into the XXth century, the adjective "Bayesian" was not introduced into the statistical lexicon until relatively recently (Fienberg 2006). It is believed that the adjective was first used in print by Ronald Aylmer Fisher in the introduction to his 1921 paper "On the probable error of a coefficient of correlation deduced from a small sample" (Fisher 1921). After the 1920s, "inverse probability" was largely replaced by a collection of methods that came to be called frequentist inference. Leonard Jimmie Savage set the stage for the neo-Bayesian revival in his book (Savage 1972), but it was the discovery of Markov chain Monte Carlo methods, which solved many of the computational problems that prevented the generalized use of Bayesian methods, that triggered a dramatic growth of the applications of these methods (Berger and Wolpert 2004).

After its revival at the end of last century, Bayesian inference has come to stay, and the number of fields in which it has been successful and usefully applied grows day by day. Just to cite some of them: biology (Huelsenbeck et al. 2002) and cosmology (Trotta 2008), in Sciences; neuroimaging (Friston et al. 2002) and hydrology (Kuczera 1999), in Engineering; and psychology (Wagenmakers et al. 2018) and economics (Koop and Korobilis 2010), in Social Sciences. The great utility and versatility of the Bayesian approach makes it a key concept not only for future scientists and engineers, but also for Social Science students. No undergraduate should graduate without at least understanding the basic concepts behind Bayesian inference (Sedlmeier 1997). Despite the importance of Bayesian inference and the growth of Bayesian research, today, most undergraduate teaching is still based on frequentist statistics; and Bayesian statistics are only introduced in some advanced graduate courses (Berry 1997).

Some of the reasons that are generally cited for explaining the absence of Bayesian statistics in undergraduate curricula are:

- The inherent difficulty of Bayesian statistics is too high for being taught at an elementary level.
- Frequentist methods still dominate in the substantive disciplines, so students must be taught these methods.
- The Bayesian approach is inherently subjective, and therefore is does not meet the objectivity standards required by Science.

On the one side, the first statement is totally wrong. In the Bayesian framework, there are only a few key basic ideas, from which all the calculations and inferences flow. The two main concepts in the Bayesian world are that uncertainties are represented by probabilities; and, of course, the Bayes' theorem itself. In contrast to the logical and intuitive interpretations of the Bayesian statistics, frequentist methods, though being relatively easy to apply, are nearly impossible to understand: in general, students in frequentist courses learn very well how to calculate confidence intervals and p-values, but they cannot give correct interpretations to these values (Berry 1997).

On the other side, the second statement is (at least for the moment) relatively right: the world of Science is still dominated by the frequentist perspective, though in the last years this is starting to change. However, even if seen as antagonists, frequentist and Bayesian frameworks are not excluding, and students can be exposed to both approaches. So, even if it is true that students should be taught frequentist methods, this does not imply that Bayesian methods have to be left out of the undergraduate curricula.

Finally, regarding the last reason of the list, Bayesian inference is, indeed, subjective; or at least, it builds a framework to incorporate subjectivity to the analysis (i.e. through the prior distribution). It is generally thought, both by educated and not so educated people, that Science is objective. Frequentists deduce from this argument that statistics must also be objective. However, the first premise is wrong (Berger and Berry 1988), as Stephen Hawking discusses in his 1988 book (Hawking 1988): Science advances with scientists modifying their opinions as new experimental data or information is gathered, and with scientists trying to convince other scientists of the correctness of their opinions.

So, the 3 main reasons why Bayesian inference is not universally taught at an undergraduate level are not such. And since the reasons why the Bayesian framework is excluded of undergraduate curricula are completely invalid, it seems reasonable to start incorporating it to undergraduate curricula as soon as possible. However, it is true that introducing Bayesian methods may be traumatic for freshmen students since learning the basics of the Bayesian approach requires developing logical thought processes. Students need both, intelligence and willingness to expend effort in thinking. And not all students meet these requirements to the same degree. The students who have the most trouble when introduced to Bayesian statistics are those who cannot unlearn the problem solving strategy developed in many high school mathematics courses, consisting in solving problems by plugging values in formulas, and where thinking is strictly optional (Berry 1997). A way of facilitating the introduction of students to the Bayesian world is to strongly reinforce the basic concepts behind the Bayesian philosophy in the first sessions devoted to this topic, in order to make sure that students completely understand the basic ideas behind this approach, rather than overwhelming them with the details of the Bayesian methods.

In this work, a simple Labview® program for reinforcing and illustrating the basic concepts underlying Bayesian inference is presented. The program consists in a computer simulation of a very basic experiment: the estimation of the bias of an unfair (or fair) virtual coin. This program may be used in a computer lab session, for the teacher to explain the concepts while the students are "playing" with the program; or as an online applet for the students to revise the concepts after the class, or in a Massive Open Online Course (MOOC) course.

2 A quick reminder: Bayesian versus frequentist approach

The main big difference between both statistics schools is the definition of probability itself. In the frequentist world, probabilities are also frequentist: probabilities represent long run frequencies. For instance, a probability of 0.4 of obtaining a "Heads" when tossing a coin, indicates that if that coin was tossed an infinite number of times (i.e. many times) the result will be "Heads" in 40% of the tosses. On the contrary, in the Bayesian world, probabilities are degrees of belief, and therefore, they can be used to represent the uncertainty in any event or hypothesis. For example, in the Bayesian approach it is totally fine to talk about the probability that "Donald Trump will win the US presidential race in 2016"; whereas a frequentist would claim that such probability is ill-defined since the event is not repeatable, and therefore, the concept of "long run frequencies" is not applicable.

Statistic inference consists in estimating a population parameter using data obtained from a sample. The difference in the probability definition between both frameworks leads to a profound nuance between frequentist inference and Bayesian inference. The first considers the parameter that is being estimated as a unknown constant parameter; and tries to find the most likely value of that parameter that is consistent with the available sample data (i.e. maximum likelihood method). On the contrary, the later considers the parameter that is being estimated as a random variable with a probability distribution that represents the uncertainty in that unknown parameter; and updates that probability distribution using the available sample data. The Cthaeh illustrates very well the difference between both approaches in his online blog PROB(A)BILISTIC WORLD (The Cthaeh 2016), where he compares the answer that would give a frequentist and the one that would give a Bayesian when faced to the question of estimating the average height of adult females:

> Frequentist
>
> "I don't know what the mean female height is. However, I know that its value is fixed (not a random one). Therefore, I cannot assign probabilities to the mean being equal to a certain value, or being less than or greater than some other value. The most I can do is collect data from a sample of the population and estimate its mean as the value which is most consistent with the data."

> Bayesian
>
> "I agree that the mean is a fixed and unknown value, but I see no problem in representing the uncertainty probabilistically. I will do so by defining a probability distribution over the possible values of the mean and use sample data to update the distribution."

(a) Heads **(b) Tails**

Fig. 1: Virtual coin

3 The experiment description

The virtual experiment considered in this work is a very simple experiment: we have a biased virtual coin, and we want to determine its bias by tossing it (i.e. simulating a toss) multiple times. As any "normal" coin, the considered virtual coin has two sides, Heads and Tails, as it can be seen in figure 1. The goal is to estimate its bias which in this context is defined as $\mathbb{P}(\text{Heads})$. By definition, this parameter is bounded between 0 and 1.

In this work the following convention will be used: random variables will be denoted by capital letters, whereas random variable values will be denoted by the corresponding lower case letter. Moreover, density functions (for continuous random variables) will be represented by f, and probability functions (for discrete random variables) will be represented by P.

In this case, the "experimental data" used to estimate the unknown parameter are the results of one (or more) tosses of the virtual coin. Y denotes the random variable "result of one toss". Since one toss only has two possible outcomes, Y is a binary discrete random variable: $y = 0$ (i.e. Tails) or $y = 1$ (i.e. Heads); and it is distributed according to a Bernoulli distribution of parameter $\mathbb{P}(\text{Heads})$.

Since this work is framed in the Bayesian approach, here the unknown parameter (i.e. the bias, $\mathbb{P}(\text{Heads})$) is considered as a random variable, X. Since the bias of a coin can be any real value between 0 and 1, X is a continuous random variable for which $x \in [0; 1]$.

The Bayes' theorem for a continuous parameter and discrete data states that:

$$f_{X|Y}(x|y) = \frac{f_X(x) \cdot P_{Y|X}(y|x)}{P_Y(y)} \propto f_X(x) \cdot P_{Y|X}(y|x) \tag{1}$$

Where $f_X(x)$ corresponds with the prior probability distribution, that carries the information (or lack of it) known before tossing the coin. $P_{Y|X}(y|x)$ is the likelihood (not to be confused with the frequentist likelihood), and gives the probability of observing y when the parameter is equal to x. Finally, $f_{X|Y}(x|y)$ corresponds with the posterior probability distribution, that carries the information known after tossing the coin. It is built by updating the prior knowledge with the experimental data (i.e. result of the toss).

On the one side, since Y is distributed according to a Bernoulli distribution of parameter x, the likelihood is given by:

$$P_{Y|X}(y|x) = x^y \cdot (1-x)^{(1-y)} \tag{2}$$

On the other side, the prior used before the first toss is selected according to the prior knowledge on the coin. For instance, in the case that no information at all is known about the coin, the definition of bias still gives some information (i.e. it is between 0 and 1). In that case a uniform prior will be used:

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Other prior can be used in case that some prior information about the coin is available. For example, if the coin is suspected to be a fair coin, a Gaussian centered around 0.5 prior should be used. On the contrary, if the coin is suspected to be an unfair coin with a tendency to Heads, a Beta picked around 0.8 should be used.

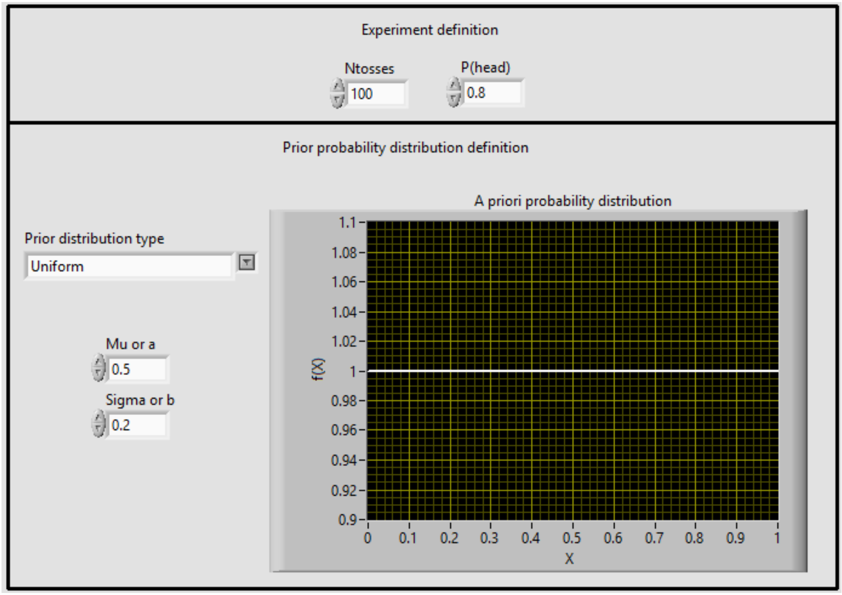
After the first toss, the initial prior will be updated using the result of the toss, obtaining the posterior distribution. This posterior distribution represents the total information about the coin after the first toss. And therefore, this distribution will be the prior distribution used for the second toss. So, in subsequent tosses, the prior is the posterior obtained after the previous toss.

4 Implementation in Labview®

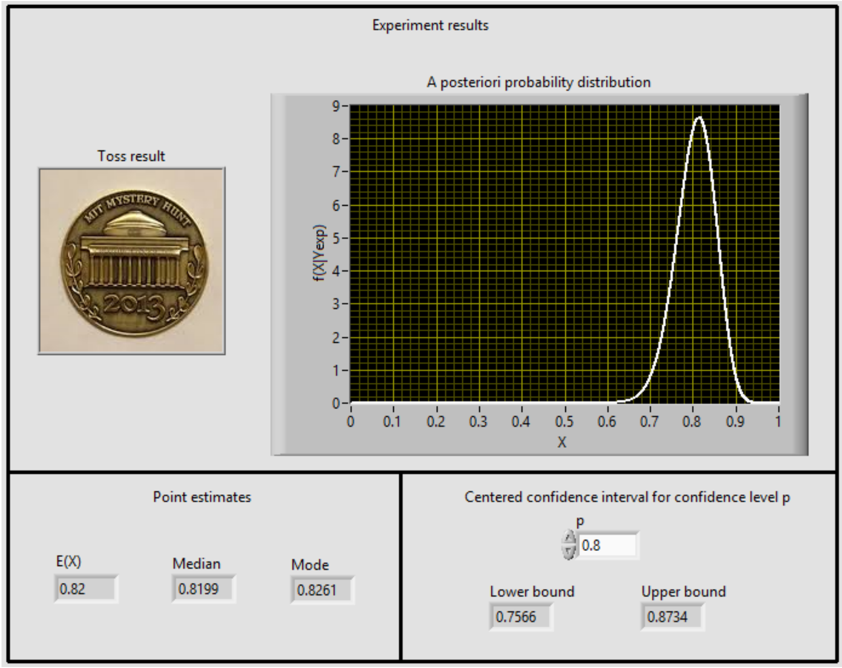
Figure 2 shows the front panel of the Labview® virtual instrument (VI) implemented in this work. It is divided in two separate blocks. On the one side, the first block is used for defining the experiment: the bias of the virtual coin, the number of tosses and the prior distribution (v.g. uniform, normal or beta). On the other side, the second block displays the results of the experiment, in real time. These results include the toss result, the calculated posterior distribution and the point estimates (i.e. mean, median, mode and probable interval). All these results are refreshed after each virtual toss.

Since Labview® cannot work with continuous variables, the continuous random variable X was discretized uniformly in its definition domain (i.e. $[0;1]$). In this context, the x-functions (i.e. $f_X(x)$, $P_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$) are represented by N_p dimensional vectors, where N_p denotes the number of discrete values of x considered in the discretization. The implemented program allows the user to select the step size, Δx. The selection of this step size should be done in order to achieve a balance between precision and computational time. All the results presented in section 5 where obtained for a step size of $1 \cdot 10^{-5}$.

The virtual coin was implemented using Labview®'s pseudo-random number generator. For each toss, the generator generates a pseudo-random number uniformly distributed between 0 and 1. If the generated number is lower than the bias of the virtual coin, selected by the user in the experiment definition block of the VI's front panel, then the result of the toss is Heads; otherwise the result is Tails.

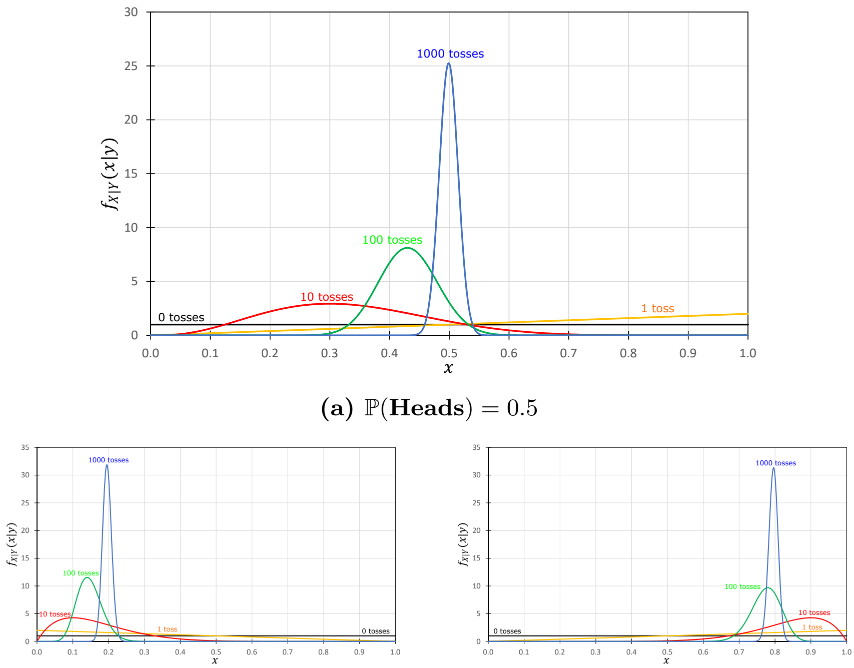


(a) Experiment definition



(b) Experiment results

Fig. 2: Labview® virtual instrument front panel



(a) $\mathbb{P}(\textbf{Heads}) = 0.5$

(b) $\mathbb{P}(\textbf{Heads}) = 0.2$

(c) $\mathbb{P}(\textbf{Heads}) = 0.8$

Fig. 3: Evolution of the posterior distribution with the number of tosses

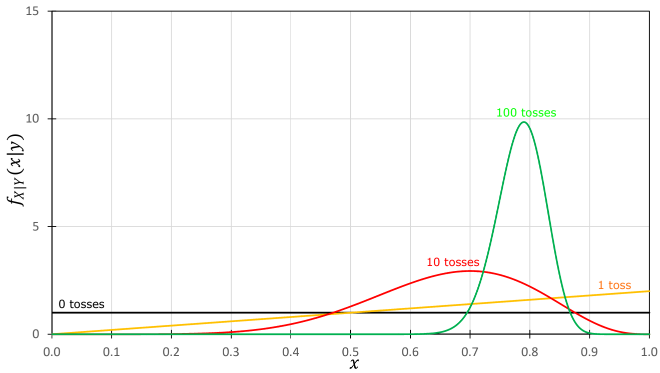
After each one of the N_t (selected by the user) tosses, the program calculates the posterior distribution by componentwise-multiplying vectors $f_X(x)$ and $P_{Y|X}(y|x)$ (equation 1). In order to avoid that numbers become too big (which would lead to an out of range error), the calculated posterior is normalized by numerically integrating the calculated vector. For the first iteration (i.e. toss), the prior selected in the experiment definition block of the VI's front panel, is used; whereas for the following iterations, the posterior of one iteration is saved to be used as the prior of the following one.

After calculating the posterior, the program calculates the point estimates. On the one hand, it calculates the mean, the median and the mode of the values stored in vector $f_{X|Y}(x|y)$. On the other hand, the program uses the posterior probability distribution, to calculate the p (selected by the user) centered confidence interval, which is the x interval that leaves $p/2$ probability tails out of the interval.

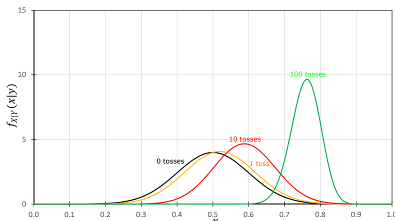
5 Result discussion

Figure 3 shows the evolution of the posterior distribution with the number of tosses starting with a uniform prior (i.e. no prior information on the coin), for 3 different virtual coins: a fair coin (i.e. $\mathbb{P}(\text{Heads}) = 0.5$), and two unfair coins ($\mathbb{P}(\text{Heads}) = 0.2$ and $\mathbb{P}(\text{Heads}) = 0.8$).

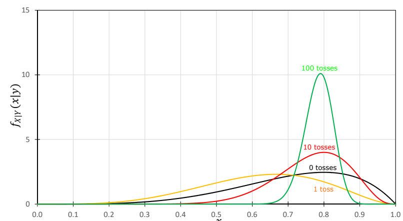
In all three cases, it can be observed that the posterior distribution narrows with each new toss, and its central position tends to the actual bias of the virtual coin. With these results, students can verify that Bayesian inference is able to correctly



(a) $\mathcal{U}(0;1)$



(b) $\mathcal{N}(0.5;0.1)$



(c) $\beta(5;2)$

Fig. 4: Evolution of the posterior distribution with the number of tosses of a $\mathbb{P}$(Heads) = 0.8 biased virtual coin, for different priors

estimate the bias of the coin, if a sufficient number of experimental data (i.e. tosses) is available. Moreover, students can visualize how getting more experimental data (i.e. more tosses) reduces the uncertainty on the parameter that is being estimated (i.e. narrower posterior distribution). Finally, watching how the posterior distribution changes in real time with each new toss, can help students understand the basic concept of Bayesian inference: new experimental data is used to update the current knowledge.

Another hot topic in the Bayesian world is how the prior selection affects the inference results; actually, the subjectivity in the prior selection is one of the main arguments that frequentists use against Bayesian inference. Students can use the Labview® program to visualize this. Figure 4 shows the evolution of the posterior distribution with the number of tosses of a unfair virtual coin with a $\mathbb{P}(\text{Heads}) = 0.8$ bias, for three different priors: a uniform prior (i.e. no prior information on the coin), a normal prior centered in 0.5 (i.e. the coin is thought to be fair), and a beta prior picked around 0.8 (i.e. the coin is suspected to the unfair and prone to Heads). And table 1 gives the point estimates obtained after 1000 tosses, en each of the aforementioned three cases.

With these results, students can clearly see how the prior selection has a great effect on the inference results for few tosses; but the effect of the prior selection dilutes out when the number of experimental data increases. They can understand that when little experimental data is available, the prior selection is very important since it has a great effect on the inference results; whereas when a lot of experimental data is available, the prior selection losses all its importance since it has no effect on the

Table 1: Point estimates after 1000 tosses of a $\mathbb{P}(Heads) = 0.8$ *biased coin, for different priors*

Point estimate	$\mathcal{U}(0;1)$	$\mathcal{N}(0.5;0.1)$	$\beta(5;2)$
Mean	0.800	0.780	0.800
Median	0.797	0.780	0.804
Mode	0.797	0.780	0.804

inference results. From this observation, students can understand the reason why all serious Bayesian inference studies include a sensitivity analysis on the prior selection (i.e. it is a way to determine whether the experimental data set is sufficiently big or not). Finally, with this example, students can see the underlying meaning behind each prior, and therefore they can understand how the prior literal information (v.g. "the coin is thought to be a fair coin") can be encoded in a probabilistic prior distribution (v.g. normal distribution centered in 0.5).

6 Conclusions

In conclusion, the Labview® program presented in this work can be used for reinforcing and illustrating the basic concepts underlying Bayesian inference: namely, the prior and posterior distributions, and how the first is updated using experimental data to get the latter. The program uses a very simple example to illustrate these concepts: the estimation of the bias of a virtual coin. By using the program students can achieve multiple outcomes, some of which are:

1. Verify that Bayesian inference is able to make accurate estimations, if a sufficient number of experimental data is available.
2. Visualize how getting more experimental data reduces the estimation's uncertainty.
3. Understand how prior information can be encoded in the prior distribution.
4. Realize the importance of the sensitivity analysis on the prior selection.

This program may be used in a computer lab session, for the teacher to explain the concepts while the students are "playing" with the program; or as an online applet for the students to revise the concepts after the class, or in a Massive Open Online Course (MOOC) course.

References

Bayes, Thomas (1763). "An essay towards solving a problem in the doctrine of chances". In: *Philosophical Transactions of the Royal Society of London* 53.3, pp. 370–418.

Berger, James O. and Donald A. Berry (1988). "Statistical analysis and the illusion of objectivity". In: *American Scientist* 76.2, pp. 159–165.

Berger, James O. and Robert L. Wolpert (2004). "A conversation with James O. Berger". In: *Statistical Science* 19.1, pp. 205–218.

Berry, Donald A. (1997). "Teaching elementary Bayesian statistics with real applications in science". In: *The American Statistician* 51.3, pp. 241–246.

De Laplace, Pierre S. (1774). "Mémoire sur la probabilité des causes par les événements". In: *Mémoires de l'Académie Royale des Sciences de Paris* 6, pp. 621–656.

Fienberg, Stephen E. (2006). "When did Bayesian inference become "Bayesian"?" In: *Bayesian Analysis* 1.1, pp. 1–40.

Fisher, Ronald A. (1921). "On the probable error of a coefficient of correlation deduced from a small sample". In: *Metron* 1, pp. 3–32.

Friston, Karl J. et al. (2002). "Classical and Bayesian inference in neuroimaging: applications". In: *Neuroimage* 16.2, pp. 484–512.

Hawking, Stephen W. (1988). *A Brief history of time: From the Big Bang to black holes*. New York: Bantam.

Huelsenbeck, John P. et al. (2002). "Potential applications and pitfalls of Bayesian inference of phylogeny". In: *Systematic Biology* 51.5, pp. 673–688.

Koop, Gary, Dimitris Korobilis, et al. (2010). "Bayesian multivariate time series methods for empirical macroeconomics". In: *Foundations and Trends in Econometrics* 3.4, pp. 267–358.

Kuczera, George (1999). "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference". In: *Water Resources Research* 35.5, pp. 1551–1557.

Savage, Leonard J. (1972). *The foundations of statistics*. New York: Courier Corporation.

Sedlmeier, Peter (1997). "BasicBayes: A tutor system for simple Bayesian inference". In: *Behavior Research Methods, Instruments, & Computers* 29.3, pp. 328–336.

Stigler, Stephen M. (1986). *The measurement of uncertainty before 1900*. Cambridge: The Belknap Press of Harvard University Press.

The Cthaeh (2016). *Frequentist and Bayesian approaches in statistics*. Online blog PROB(A)BILISTIC WORLD. URL: https://www.probabilisticworld.com/frequentist-bayesian-approaches-inferential-statistics/.

Trotta, Roberto (2008). "Bayes in the sky: Bayesian inference and model selection in cosmology". In: *Contemporary Physics* 49.2, pp. 71–104.

Wagenmakers, Eric J. et al. (2018). “Bayesian inference for psychology. Part II: Example applications with JASP”. In: *Psychonomic Bulletin & review* 25.1, pp. 58–76.