International Journal of Cardiology 165 (2013) 3–6

Contents lists available at SciVerse ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard

Editorial

A 64,489-patient full-disclosure database of cardiovascular risk factors and events status analysed in a Bayesian framework: A unique contribution to predictive science

Hakim-Moulay Dehbi *, Darrel P. Francis

International Centre for Circulatory Health, Imperial College London, London, UK

ARTICLE INFO

Article history:
Received 10 September 2012
Accepted 26 September 2012
Available online 20 November 2012

Keywords:
Cardiovascular risk factors
Events status
Predictive distribution
Bayesian clinical reasoning
Bayesian theorem

ABSTRACT

Today in the International Journal of Cardiology Liu et al. [1] publish an unusual exercise in open science which should set a pioneering trend for future knowledge sharing. They present both the principle and a large fully-analysed real world dataset to show how Bayesian reasoning can be practically helpful for clinicians at the front line.

The Bayesian approach differs from the frequentist approach that is more commonly seen in reports of clinical research. Instead of a probability having a single point estimate and confidence interval, it instead has a complete probability density function. For Bayesian analysis in general, instead of there being no information before a particular study, there is some information — the "prior". The difference is that while the frequentist approach assumes that before the study all probabilities are equally plausible, the Bayesian approach recognises that even before the study, some probabilities are more likely than others. Therefore, after the study, the Bayesian approach produces a new distribution of the probability – the "posterior" – which incorporates both the raw study results and the prior distribution.

Bayesian approaches are routinely used in medical decision-making and everyday life, perhaps without even realising it. Clinical test results are rarely interpreted in isolation. Instead, the background clinical belief of plausibility of various diagnoses (the prior) is updated in light of test results, to form a new set of beliefs (the posterior). We more readily accept assertions that are within the range of our prior beliefs than those that substantially contradict those beliefs.

To build a model of cardiovascular risk, the Bayesian approach begins with an assumed distribution for the risk depending on the risk factors and progressively updates it with the experience of patients and their outcomes. Each additional patient makes a contribution to the model's knowledge. Then the model can be applied to any individual, and provide a distribution for the risk of that individual. This might be narrow, indicating precise risk evaluation or wide, indicating substantial persisting uncertainty.

The authors' openness to share the whole dataset creates three exciting avenues for advancement in the field. First, researchers could analyse the dataset in different ways, for example, by proposing distributions other than the normal. Second, they could use the outcome of this dataset as a starting point for further upgrading the model with future data. Third, researchers are absolutely free to use this data to explore other interrelationships between the variables for new purposes. For example, we have studied the joint distribution of two variables that have a multiplicative effect of cardiovascular risk: cholesterol and blood pressure. The online supplement to this editorial contains the raw dataset in .zip format to facilitate its download for the reader.

Freely exposing all the data is currently remarkable but on objective reflection, it is hard to understand why it is not already a normal practice. Do authors fear that readers despite a handicap of years might trump them to future findings? Or do they have something to hide? We do not know but this paper is changing our practice and we hope it will change yours.

© 2012 Elsevier Ireland Ltd. All rights reserved.

An open mind is an empty mind.

Bertrand Russell

The central principle of Bayesian statistics is that our knowledge about anything can be described by a probability density function,

* Corresponding author at: International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, 59-61 North Wharf Road, London W2 1LA, UK. Tel.: +44 7949252915.
E-mail address: h.dehbi@imperial.ac.uk (H.-M. Dehbi).

0167-5273/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijcard.2012.09.209

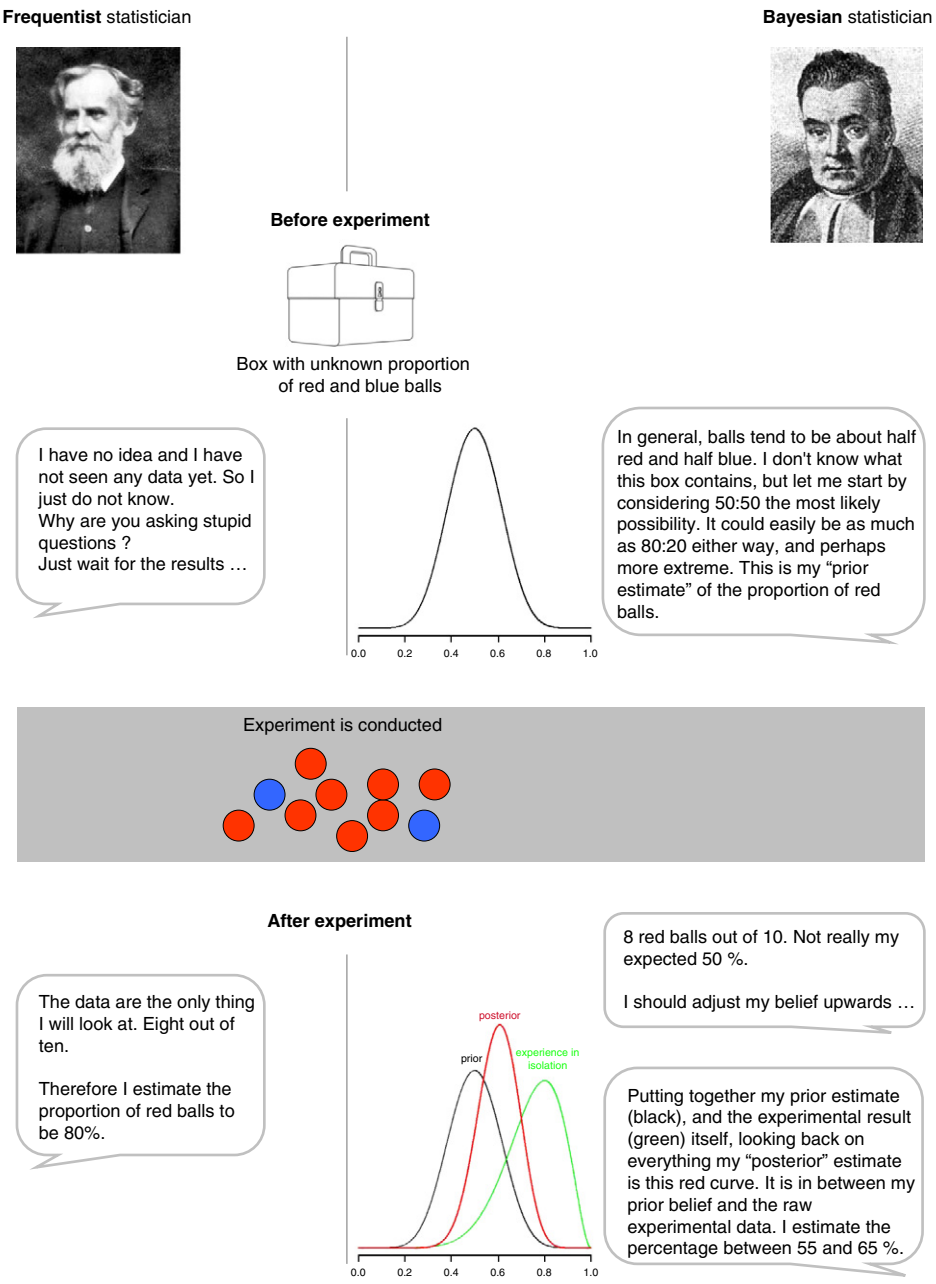

Fig. 1. Illustration of the difference between Bayesian and frequentist approaches. In this thought experiment a Bayesian statistician and a frequentist statistician separately estimate the proportion of red balls in a large box containing numerous balls, some red and some blue. They do not know the real proportions, which are 60% red and 40% blue, and ask a neutral person to ten times mix the balls, draw out one ball, report its colour, and then replace it in the box. Out of these ten drawings, two are blue and eight are red.

with some possibilities more plausible than others, and that this can be serially revised as new information arrives (Fig. 1). In this issue, Liu et al. [1] use Bayes' theorem to predict the risk of cardio-vascular disease (CVD) using a dataset of 64,489 patients coming from a community-based screening programme in Taiwan. By the central Bayesian principle, the CVD risk in this study is represented by a probability density function. This constitutes a fundamental difference with classical statistics where the CVD risk is considered as fixed but unknown. Indeed whilst in classical statistics the data are considered to provide information about the unknown risk, in Bayesian statistics the data are used to update one's a-priori knowledge about the risk. Bayes' theorem accommodates the a-priori knowledge with a prior distribution and adjusts it with the likelihood of the data. This process results in a posterior distribution that incorporates both the initial knowledge and the information coming from the data (Fig. 2).

Both the prior distribution of the CVD risk and the likelihood were assumed to be normal. Using sequential learning, the model was refined and upgraded to take into account additional information coming from the experience of patients. Initially, the model included the variables gender, age, age^2 and period of recruitment as predictors to form the likelihood. The posterior obtained was used as a prior distribution for the CVD risk of the upgraded model containing additionally information on six metabolic syndrome components. Finally, this second posterior was exploited as prior for the final model incorporating information on smoking, drinking, betel-chewing and family-history of CVD. Using simulations, the third posterior obtained was employed to derive individual predictive distributions based on the risk factors. This process called Bayesian clinical reasoning gives CVD risk distributions corresponding to the various possible combinations of the risk factors, which can be practically helpful for clinicians at the front line.

Download English Version:

https://daneshyari.com/en/article/5976984

Download Persian Version:

https://daneshyari.com/article/5976984

Daneshyari.com