Book of Short Papers SIS 2020

Editors: Alessio Pollice, Nicola Salvati and Francesco Schirripa Spagnolo

Published by Pearson

www.pearson.com

ISBN 9788891910776

Contents

Specialized sessions

Solicited Sessions

Contributed papers and Posters

This book is our contribution to encourage the scientific community and the network of the Italian Statistical Society to go on and transform this difficult period into an opportunity of scientific debate for better statistics in a better world.

Alessio Pollice
Università degli Studi di Bari Aldo Moro
Chair of the Program Committee

Nicola Salvati
Università di Pisa
Chair of the Local Organizing Committee

Francesco Schirripa Spagnolo
Università di Pisa

Program Committee: Alessio Pollice (Chair), Serena Arima, Marilena Barbieri, Alessandra Brazzale, Eugenio Brentari, Alessia Caponera, Antonio Lepore, Antonella Plaia, Tommaso Proietti, MarcoRiani, Nicola Salvati, Pasquale Sarnacchiaro, Mauro Scanu, Manuela Stranges, Valentina Tocchioni, Simone Vantini, Massimo Ventrucci, Paola Vicard, Donatella Vicari.

Local Organizing Committee: Nicola Salvati (Chair), Gaia Bertarelli, Bruno Cheli, Alessandra Coli, Paolo Frumento, Fosca Giannotti, Caterina Giusti, Piero Manfredi, Stefano Marchetti, Lucio Masserini, Vincenzo Mauro, Barbara Pacini, Dino Pedreschi, Francesco Schirripa Spagnolo, Chiara Seghieri.

Organizers of Specialized and Solicited Sessions: Giada Adelfio, Bruno Arpino, Emanuele Aliverti, Nicoletta Balbo, Mara Bernardi, Silvia Bozza, Pierpaolo Brutti. Annalisa Busetta, Michela Cameletti, Carlo Cavicchia, Fabrizio Durante, Leonardo Egidi, Pietro D. Falorsi, Francesco Finazzi, Livio Finos, Stefania Galimberti, Michele Gallo, Caterina Giusti, Francesca Greselin, Alessandra Guglielmi, Francesca Ieva, Tiziana Laureti, Achille Lemmi, Brunero Liseo, Fabio Massimo Lo Verde, Daria Mendola, Roberta Pappadà, Lea Petrella, Alessandra Petrucci, Alessia Pini, Sabrina Prati, Maria Giovanna Ranalli, Davide Risso, Fabrizio Ruggeri, Silvana Salvini, Monica Scannapieco, Francesco Stingo, Luca Tardella, Grazia Vicario, Susanna Zaccarin, Maroussa Zagoraiou.

Wasserstein consensus for Bayesian sample size determination

Consenso di Wasserstein per la determinazione Bayesiana della dimensione campionaria

Michele Cianfriglia, Tullia Padellini and Pierpaolo Brutti

Abstract The sample size determination problem deals with the selection of the optimal number of subjects to be enrolled in a study in order to achieve a pre–specified inferential goal. While this problem can of course be approached from a frequentist viewpoint, often the Bayesian paradigm is preferred as it allows to blend and balance the strength of the observed empirical evidence with the available prior knowledge. In this work, we focus on the case of a "community of priors" representing, for example, different expert opinions. Within this setup, we are interested in selecting the smallest sample size that guarantees "agreement" between these, possibly conflicting, opinions, having formalized the loose idea of "agreement" in terms of the Wasserstein distance between posteriors stemming from different priors.
Abstract *Il problema della determinazione della dimensione campionaria concerne la selezione del numero ottimo di soggetti da considerare in uno studio al fine di raggiungere un prestabilito livello di accuratezza inferenziale. In quest'ambito, l'approccio Bayesiano è in genere preferito perché permette di bilanciare i dati osservati con le informazioni a priori disponibili. Scopo di questo lavoro è partire da una "famiglia di a priori" che rappresentano, ad esempio, le opinioni di un pool di esperti, ed arrivare ad individuare il numero minimo di osservazioni necessario a garantire un accordo tra le diverse posizioni considerate. Il concetto di accordo è formalizzato attraverso la distanza di Wasserstein tra le distribuzioni a posteriori ottenute, ad esempio, a partire dalle diverse opinioni a priori degli esperti coinvolti.*

Key words: Bayesian consensus, sample size determination, Wasserstein distance, clinical trials.

Michele Cianfriglia · Pierpaolo Brutti
Dipartimento di Scienze Statistiche, Sapienza Università di Roma
e-mail: {michele.cianfriglia}, {pierpaolo.brutti}@uniroma1.it

Tullia Padellini
Department of Epidemiology and Biostatistics, Imperial College London
e-mail: t.padellini@imperial.ac.uk

1 Introduction

In the design of a statistical experiment a fundamental issue is the determination of the optimal sample size (SSD). Focusing, for example, on clinical trials, SSD is particularly critical as there is an immediate interpretation of the optimal sample size in terms of the minimum number of subjects to be enrolled in order to achieve a pre–specified inferential target. Historically, while both frequentist and Bayesian solutions have been provided for this problem, the Bayesian paradigm is often preferred as it allows to combine observed empirical evidence with all the prior knowledge already available, including historical data. Among others, examples of this approach can be found in [3, 7, 8, 10] and [4].

In their general formulation, common Bayesian criteria used to determine the optimal sample size need to control the inferential performance of specific functionals of interest of the posterior distribution, e.g. posterior mean or credible interval. For this reason the optimal sample size is usually defined as the minimum sample size for which the aforementioned functionals achieve some inferential goal on average or with large enough (predictive) probability.

Our goal in this work is to address the SSD problem in case where we are dealing with more than one prior distribution for the parameter of interest, a "community of priors". Indeed this is a not an uncommon setting as it can be used, for example, to model multiple scenarios we want to account for, or even different expert opinions. A possible approach to this problem detailed in [2], consists in aggregating different beliefs through the use of finite mixture models. Here however we adopt a different definition of consensus, which is not based on aggregation but rather on *distance minimization*. To this end, our work heavily relies on the *Wasserstein metric*, a well known distance between probability distributions that nowadays is gaining momentum in the statistics and machine learning communities.

Hence, once the loose idea of "agreement" has been formalized in terms of the Wasserstein distance between posteriors stemming from different prior distributions, we propose a Bayesian SSD procedure for the selection of the smallest sample size that guarantees "agreement" between possibly conflicting prior opinions. We detail our proposal in the case of the ubiquitous conjugate *Normal model* for which the Wasserstein distance admits an exact expression neatly depending on the posterior means and standard deviations of the distributions involved.

This paper is organized as follows. In the next section we introduce the new definition of consensus. In Section 3 two main Bayesian sample size criteria, that is the Predictive Expectation Criterion and the Predictive Probability Criterion, are redefined in terms of the Wasserstein distance, and an application to real data is reported. Finally Section 4 shows some motivating results based on the data analyzed.

2 A new perspective of consensus: Wasserstein consensus

In designing a new clinical trial it is quite common to take into account the results provided by previous, related, medical studies or to consult several experts. Since the information provided by these variety of sources can happen to be quite conflicting, once the uncertainty surrounding them is captured in terms of a suitable probability distribution, we are left with the statistical problem of working with a "community of priors" elicited over the same parameter of interest.

In the literature, conflicting prior opinions are typically handled by *robust methods* – such as finite mixture or ε–contaminated priors – as, for example, in [1, 2, 4]. There are two main issues with this approach: (i) we lose in "resolution" since we start the Bayesian machinery from a single, encompassing, prior built to cover the range, from optimistic to pessimistic, spanned by the pool of elicited distributions; and consequently (ii) we also lose the opportunity to address the SSD problem, or any other design problem for what matters, with the explicit goal of achieving any sort of "agreement" between conclusions stemming from different scenarios.

Broadly speaking, there are of course many ways to define "agreement". Within the Bayesian framework we are working with, this loose idea naturally translates into guaranteeing *posterior* "agreement" between possibly conflicting prior opinions. To date, the only work which addresses the issue of determining the optimal sample size in order to guarantee posterior consensus is [6].

In this work, we formalize Bayesian consensus via the Wasserstein distance between posterior distributions stemming from different priors. To this end, taking inspiration from the main Bayesian sample size criteria, we want to find the minimum number of subjects which ensures that a suitable posterior summary of the Wasserstein distance is lower than a pre–specified threshold.
The use of the Wasserstein distance has several advantages, for instance:

- we easily define "agreement" between experts: two experts "agree" if their inferential conclusions, namely their posterior distributions, are "close" enough;
- it "metricizes" convergence in distribution: if two distinct distributions are close with respect to the Wasserstein distance, then they are probabilistically similar;
- it tells us why distributions differ: the Wasserstein distance is a transportation distance and comes with a transportation plan that detail how to move the mass of a distribution to morph it into another.

3 Our proposal: Wasserstein based criteria

Before describing our proposal, let us recall the basic ideas and definitions behind the two main Bayesian sample size criteria. Generally speaking, these criteria are defined in terms of suitable summaries of the posterior distribution, that is

$$\rho_\pi(\theta|\mathbf{x}_n) = \int g(\theta)\pi(\theta|\mathbf{x}_n)\,\mathrm{d}\theta,$$

where different choices of $g(\theta)$ lead to different summaries of $\pi(\theta|\mathbf{x}_n)$. Denoting by $m(\cdot)$ the prior predictive distribution, we are now in position to define

- *Predictive Expectation Criterion (PEC)*: let e_n be the expected value of $\rho_\pi(\theta|\mathbf{x}_n)$ with respect to $m(\cdot)$. Given a suitable threshold η_e the optimal sample size is
$$n_e^\star = \min\{n \in \mathbb{N} : e_n > \eta_e\};$$
- *Predictive Probability Criterion (PPC)*: let p_n be the probability with respect to $m(\cdot)$ that $\rho_\pi(\theta|\mathbf{x}_n)$ is bigger than a constant γ. Given a suitable threshold $\eta_p \in (0,1)$ the optimal sample size is
$$n_p^\star = \min\{n \in \mathbb{N} : p_n > \eta_p\}.$$

A few comments: (i) these criteria are called *predictive* because they use the marginal predictive distribution $m(\cdot)$; (ii) in defining the Bayesian criteria we usually consider two distinct prior distributions, specifically π_A (*analysis prior*) – which is employed in determining the posterior distribution – and π_D (*design prior*) – which induces the marginal distribution of the data.

We can now define the Wasserstein based criteria. In the following, d_W denotes the 2–Wasserstein distance, based on the L_2 ground metric, between two posterior distributions derived by two different analysis priors.

- *Predictive Expectation Criterion (PEC)*: let e_n^W be the expected value of d_W with respect to $m(\cdot)$. Given a suitable threshold η_e the optimal sample size is
$$n_{e,W}^\star = \min\{n \in \mathbb{N} : e_n^W < \eta_e\}; \tag{1}$$
- *Predictive Probability Criterion (PPC)*: let p_n^W be the probability with respect to $m(\cdot)$ that d_W is bigger than a constant γ. Given a suitable threshold $\eta_p \in (0,1)$ the optimal sample size is
$$n_{p,W}^\star = \min\{n \in \mathbb{N} : p_n^W < \eta_p\}. \tag{2}$$

Please notice that the optimal sample sizes based on these criteria have a *particular* interpretation because we are searching for the smallest sample size that guarantees "agreement" between possibly conflicting opinions, therefore it is natural to reverse the inequality of the Bayesian criteria as done in (1) and (2). In addition, under a Normal conjugate model, the functions e_n^W and p_n^W admit a closed form expression making the computation of the Wasserstein based criteria extremely efficient. This is quite crucial since the computation of the Wasserstein distance becomes quite expensive and time consuming when multidimensional posterior distributions are involved.

4 A real data example

We conclude by showing how the Wasserstein based–criteria perform on a real data application. More in details, we revisit an example reported in [9], where results of a meta–analysis are reinterpreted from a Bayesian perspective. The setup is the following: a series of small randomized trials was conducted to assess a proactive effect of intravenous magnesium sulphate after acute myocardial infarction. Even if many studies were already conducted, further investigation was suggested and the massive ISIS-4 trial started. Unfortunately, this study did not provide evidence of any benefit contradicting previous conclusions (see [5] and Figure 1 for details).

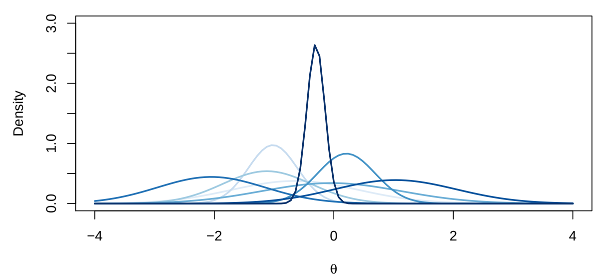

Fig. 1 Comparison between Normally distributed priors encoding evidence from previous studies. The parameter of interest θ is the log odds ratio of intravenous magnesium with respect to placebo. As suggested in [9], the variance of the analysis prior is σ^2/n_0, where $\sigma^2 = 4$ is fixed and n_0 represents the *effective number of events* in the study considered.

Table 1 shows the optimal sample size $n^\star_{e,W}$ obtained using our version of the PEC with $\eta_e = 0.05$. The marginal predictive distribution is based on a Normally distributed design prior with mean 0.058 and variance $\sigma^2 = 4/n_D$. Intuitively, different choices of the design prior sample size n_D reflect different strength of the skepticism toward the magnesium treatment: as n_D increases it becomes easier and easier to bring consensus between different parties since we have more and more information about the phenomenon. From a mathematical point of view, this aspect can be understood looking at the behavior of e^W_n as function of n_D.

n_D	$n^\star_{e,W}$	$n^\star_{e,M}$
4319	361	498
432	371	509
43	468	190

Table 1 Comparison between optimal sample sizes. The first column contains the prior sample size of the design prior, whereas the other two columns contain, respectively, optimal sample sizes associated with the Wasserstein and mixture versions of the PEC.

Finally, in order to highlight the need for “consensus–assessment”, we compare our optimal sample sizes with those obtained by [2], where a finite mixture version of the standard PEC is used. Note that the two methods are *not* directly comparable as their inferential goals are different.

Nevertheless we stress the fact that consensus does not come automatically and “for free”: the more the parties disagree, the higher is the number of subjects to be enrolled in the study in order to achieve inferential agreement.

References

1. Brutti, P., De Santis, F., Gubbiotti, S.: Robust bayesian sample size determination in clinical trials. Statistics in Medicine **27**(13), 2290–2306 (2008)
2. Brutti, P., De Santis, F., Gubbiotti, S.: Mixtures of prior distributions for predictive bayesian sample size calculations in clinical trials. Statistics in medicine **28**(17), 2185–2201 (2009)
3. Brutti, P., De Santis, F., Gubbiotti, S.: Bayesian–frequentist sample size determination: a game of two priors. Metron **72**(2), 133–151 (2014)
4. De Santis, F.: Using historical data for bayesian sample size determination. Journal of the Royal Statistical Society: Series A (Statistics in Society) **170**(1), 95–113 (2007)
5. Higgins, J.P., Spiegelhalter, D.J.: Being sceptical about meta-analyses: a bayesian perspective on magnesium trials in myocardial infarction. International journal of epidemiology **31**(1), 96–104 (2002)
6. Joseph, L., Bélisle, P.: Bayesian consensus-based sample size criteria for binomial proportions. Statistics in medicine (2019)
7. Joseph, L., Bélisle, P., Du Berger, R.: Bayesian and mixed bayesian/likelihood criteria for sample size determination. Statistics in medicine **16**(7), 769–781 (1997)
8. Joseph, L., Du Berger, R., Wolfson, D.B.: Some comments on bayesian sample size determination. Journal of the Royal Statistical Society: Series D (The Statistician) **44**(2), 167–171 (1995)
9. Spiegelhalter, D.J., Abrams, K.R., Myles, J.P.: Bayesian approaches to clinical trials and health-care evaluation, vol. 13. John Wiley & Sons (2004)
10. Wang, F., Gelfand, A.E., et al.: A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. Statistical Science **17**(2), 193–208 (2002)