Scientific Research Publishing

Open Journal of Applied Sciences, 2025, 15(9), 2545-2555
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

On the Interpretation and the Choice of the Hyperparameters of a Beta Prior Distribution

Valeria Sambucini

Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
Email: valeria.sambucini@uniroma1.it

How to cite this paper: Sambucini, V. (2025) On the Interpretation and the Choice of the Hyperparameters of a Beta Prior Distribution. *Open Journal of Applied Sciences*, **15**, 2545-2555.
https://doi.org/10.4236/ojapps.2025.159169

Received: June 25, 2025
Accepted: August 31, 2025
Published: September 3, 2025

Abstract

Binary outcomes are frequently encountered in a variety of fields and contexts and the Bayesian approach is widely used to analyze this type of data. Under this framework, a beta prior distribution for the probability of success is typically used. We clarify why in the statistical literature one can find two slightly different interpretations of the prior hyperparameters as extra data. The two interpretations can be exploited to elicit informative beta prior densities by specifying a measure of central tendency and a suitable value for the prior sample size. We developed a Shiny App in R that provides a user-friendly interface to implement the elicitation procedures.

Keywords

Beta Prior Distributions, Binary Data, Elicitation, Extra Data, Prior Sample Size

1. Introduction

Binary outcomes, which assume only two possible values, are often analysed in a variety of fields and contexts, especially in epidemiology and clinical trials. In fact, many standard epidemiological problems, such as the study of the effectiveness of a new vaccine or the evaluation of diagnostic tests and the assessment of their performance, are based on binary variables. Cohort and case-control studies, conducted to identify aetiological agents that increase the risk of a certain disease, are typically based on a binary exposure variable. Also in clinical trials, the primary endpoint is often a binary endpoint, that allows each patient to be classified as a responder or not to the experimental drug: standard examples are single-arm

phase II trials or comparative phase III studies that exploit the risk difference, the relative risk or the odds-ratio as parameters of interest. In all these areas of research, the Bayesian approach is widely used to analyze binary outcomes.

The Bayesian analysis of binary data, generated from a Bernoulli distribution, requires the specification of a prior distribution for the unknown parameter θ, that represents a success probability. A very common choice in this situation is the beta distribution. In addition to computational convenience and high flexibility, this distribution allows a nice interpretation of its hyperparameters as extra data. However, there are two possible and slightly different interpretations. According to the first one, the information contained in a beta prior density of hyperparameters α and β corresponds to the augmentation of data with α successes and β failures [1]-[4]. The alternative interpretation suggests that the amount of data to add to the current experiment consists in $\alpha-1$ successes and $\beta-1$ failures [5]-[8]. Both the interpretations are mathematically sound, but they reflect different philosophical stances on prior information strength. Moreover, most textbooks and papers from the statistical literature typically provide only one of them, which can lead to ambiguity and confusion.

The rest of the paper is organized as follows. In Section 2, we revise the Bayesian analysis of binary data. In Section 3, we clarify why both the interpretations of the hyperparameters are reasonable by illustrating the reasoning that supports them. Section 4 shows how the two interpretations can be exploited to elicit informative beta prior densities. In Section 5, we present a Shiny App in R that provides a user-friendly interface to implement the elicitation procedures and to help students or practitioners not experts in Bayesian statistics to understand the role of the prior sample size. The main functionalities of the app are illustrated through numerical examples. Finally, Section 6 contains some concluding remarks.

2. Bayesian Analysis of Binary Data

Let us introduce an experiment based on a binary response variable Y related to an event's occurrence. Thus, we consider a random sample, $(Y_1,\cdots,Y_n)$, with independent and identically distributed variables such that $Y_i \mid \theta \sim Bernoulli(\theta)$, for $i=1,\cdots,n$, where θ is the unknown success probability, *i.e.* the probability that the event occurs. To summarize data, we typically use the sum of successes, that is the sufficient statistic $S_n=\sum_{i=1}^{n} Y_i$ with a binomial sampling distribution of parameters (n,θ). Under a frequentist framework, the maximum likelihood estimate of θ is the sample mean, $\hat{\theta}_{ML}=s_n/n$, *i.e.* the frequency of the successes among all observations.

Under the Bayesian approach, before data is observed, the parameter θ is treated as a random variable having a prior distribution, $\pi(\theta)$, that expresses pre-experimental knowledge and belief available about θ. When dealing with binary data, a beta prior distribution for θ is commonly used, that is

$$\pi(\theta)=Beta(\theta;\alpha,\beta)=\frac{1}{\mathcal{B}(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},$$

where the hyperparameters α and β are both larger than 0 and $\mathcal{B}(\alpha,\beta)$ denotes the beta function. The mode, the expected value and the variance of $\pi(\theta)$ are, respectively,

$$Mode(\theta)=\frac{\alpha-1}{\alpha+\beta-2},\ E(\theta)=\frac{\alpha}{\alpha+\beta}\ \text{and}$$

$$Var(\theta)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

This choice is mainly due to analytical tractability, since the class of the beta prior densities is conjugate to the binomial model and, therefore, the corresponding posterior distribution will belong to the same family of distributions as the prior. A further reason that motivates the use of a beta prior is that its probability density function can take many different shapes depending on the two hyperparameters and, thus, it can be exploited to model very different kinds of information.

Assuming that s_n successes have been observed, from standard results of conjugate analysis, the posterior distribution of θ is still a beta density with updated parameters [5], that is

$$\pi(\theta\mid s_n)=Beta(\theta;s_n+\alpha,n-s_n+\beta).$$

As it is well known, Bayesian inference about θ is realized by summarizing the information contained in its posterior distribution. For instance, a Bayesian point estimate of θ can be obtained by considering a measure of central tendency of $\pi(\theta\mid s_n)$. The most commonly used measures are the posterior mode or the posterior mean. By considering one or the other of these measures, we can derive the two interpretations of the hyperparameters of $\pi(\theta)$ mentioned above.

3. Interpretation of the Hyperparameters of the Beta Prior Distribution

Let us start by considering the posterior mode as a Bayesian estimate of θ,

$$\begin{aligned}Mode(\theta\mid s_n)&=\frac{\alpha+s_n-1}{\alpha+\beta+n-2}\\&=\frac{\alpha-1}{\alpha+\beta-2+n}+\frac{s_n}{\alpha+\beta-2+n}.\end{aligned}\tag{1}$$

By multiplying the first term of (1) by $\frac{\alpha+\beta-2}{\alpha+\beta-2}$ and by making the substitution $s_n=n\hat{\theta}_{ML}$ in the second term, we obtain that

$$\begin{aligned}Mode(\theta\mid s_n)&=\frac{\alpha+\beta-2}{\alpha+\beta-2+n}Mode(\theta)+\frac{n}{\alpha+\beta-2+n}\hat{\theta}_{ML}\\&=\frac{n_0^M}{n_0^M+n}Mode(\theta)+\frac{n}{n_0^M+n}\hat{\theta}_{ML},\end{aligned}$$

where $n_0^M=\alpha+\beta-2$. In practice, the mode of the posterior distribution, which represents the most plausible value a posteriori, is a weighted average of the prior

mode and the sample estimate of θ based on the observed data. This result clearly shows how the beta posterior distribution combines prior beliefs about θ with observed data and turns out to be an updated probability distribution. The weights assigned to $\hat{\theta}_{ML}$ and to the prior mode are, respectively, the sample size of the experiment n and $n_0^M = \alpha + \beta - 2$, that can be interpreted as the *prior sample size*. The result, in fact, also suggests that the information provided by the beta prior density is equivalent to that of a virtual binomial experiment based on $\alpha + \beta - 2$ observations, with $\alpha - 1$ successes and $\beta - 1$ failures. The correspondence between the observed number of successes in the current experiment and the virtual number of prior successes is evident from (1).

A similar, but slightly different, interpretation of the hyperparameters of the beta prior arises if we consider as Bayesian estimate of θ the posterior mean. In this case, we have that

$$E(\theta \mid s_n) = \frac{\alpha + s_n}{\alpha + \beta + n} = \frac{\alpha}{\alpha + \beta + n} + \frac{s_n}{\alpha + \beta + n}.$$

With computations analogous to those performed for the posterior mode, we obtain that

$$\begin{aligned} E(\theta \mid s_n) &= \frac{\alpha + \beta}{\alpha + \beta + n} E(\theta) + \frac{n}{\alpha + \beta + n} \hat{\theta}_{ML} \\ &= \frac{n_0^E}{n_0^E + n} E(\theta) + \frac{n}{n_0^E + n} \hat{\theta}_{ML}, \end{aligned}$$

where $n_0^E = \alpha + \beta$. Thus, the mean of the posterior distribution turns out to be a weighted average of the prior mean and the ML estimate of θ. The contribution of $\hat{\theta}_{ML}$ is again the sample size n, while the weight of the prior mean is $n_0^E = \alpha + \beta$, that plays the role of the *prior sample size*. Hence, in this case, the beta prior distribution corresponds to adding α prior successes and β prior failures.

Both the interpretations of the beta prior density are equally legitimate based on the aforementioned arguments. Note that the non-informative prior distribution corresponding to a sample with no observations is the uniform distribution $Beta(1,1)$ under the interpretation based on the posterior mode and the improper Haldane's distribution $Beta(0,0)$ under the interpretation based on the posterior mean.

4. Elicitation of the Prior Hyperparameters

The two interpretations of the hyperparameters of the prior $\pi(\theta) = Beta(\theta; \alpha, \beta)$ as extra data can be exploited to construct an informative prior distribution for θ based on pre-experimental information.

The idea is to express the hyperparameters in terms of 1) a measure of central tendency of the prior and 2) the prior sample size. More specifically, if we opt for the prior mode θ_0^M, we need to consider the following system of equations

$$\begin{cases} \theta_0^M = \dfrac{\alpha - 1}{\alpha + \beta - 2}, \\ n_0^M = \alpha + \beta - 2 \end{cases}$$

obtaining the solutions

$$\alpha = n_0^M \theta_0^M + 1 \text{ and } \beta = n_0^M \left(1 - \theta_0^M\right) + 1. \tag{2}$$

Thus, this choice of the hyperparameters ensures that the mode of the beta prior is θ_0^M, that can be selected as the value of the parameter considered the most plausible according to the prior information. Then, we can regulate the concentration of $\pi(\theta)$ around θ_0^M through the selection of the prior sample size n_0^M. This quantity, in fact, represents the number of virtual observations to which the prior information is considered equivalent and, therefore, the larger n_0^M, the more concentrated the prior distribution. This can also be easily proved by noting that the variance of the beta prior distribution with hyperparameters provided in (2), that is

$$\begin{aligned} Var(\theta) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{\left(n_0^M \theta_0^M + 1\right)\left(n_0^M - n_0^M \theta_0^M + 1\right)}{\left(n_0^M + 2\right)^2 \left(n_0^M + 3\right)}, \end{aligned}$$

decreases as n_0^M increases. Moreover, as $n_0^M \to \infty$, the limit of the variance is 0 and, thus, if n_0^M tends to infinity the beta prior density tends to a degenerate distribution at θ_0^M. This elicitation procedure has been exploited, for instance, by [9] [10] at the design stage of a clinical trial and it is often applied by setting the prior sample size to 1, in order to obtain a weakly informative prior [11] [12].

Analogously, if we choose the prior mean θ_0^E as measure of central tendency, by solving the following system of equations

$$\begin{cases} \theta_0^E = \dfrac{\alpha}{\alpha + \beta}, \\ n_0^E = \alpha + \beta \end{cases}$$

we obtain that

$$\alpha = n_0^E \theta_0^E \text{ and } \beta = n_0^E \left(1 - \theta_0^E\right) \tag{3}$$

are the hyperparameters of a beta prior density with mean equal to θ_0^E and prior sample size n_0^E. Also in this case, as n_0^E increases, the variability of the beta prior decreases and the distribution turns out to be more concentrated. In the limiting case, when $n_0^E \to \infty$, $\pi(\theta)$ corresponds to the probability distribution that assigns all the probability mass to θ_0^E.

In **Figure 1**, we compare the beta prior distributions obtained by specifying the prior mode and the prior mean, assuming equal values for these measures of central tendency and for the prior sample size. As expected, the prior densities tend to coincide as the central tendency approaches 0.5 and as $n_0^M = n_0^E$ increases.

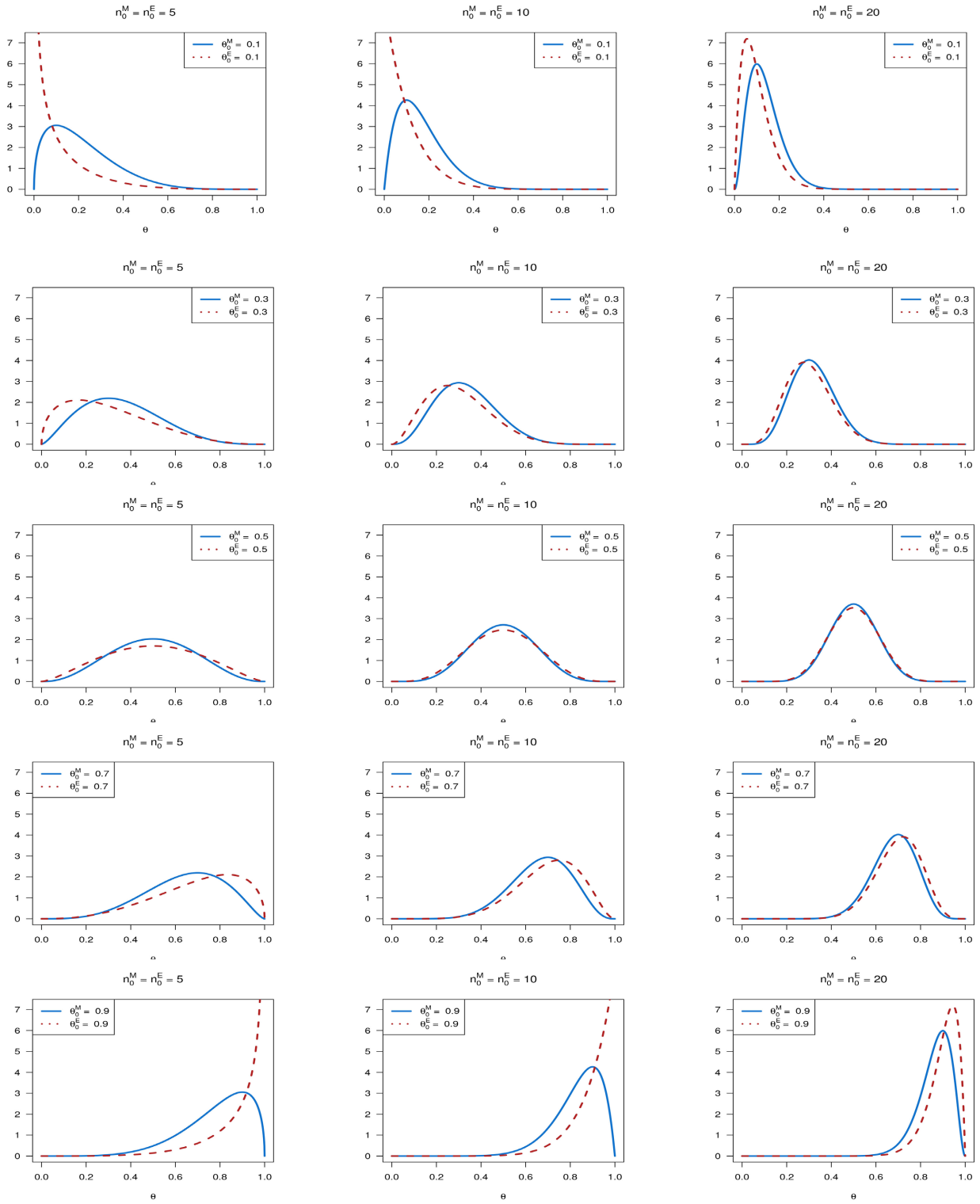

Figure 1. Comparison of beta prior distributions for θ when using the hyperparameters in (2) and (3) for different values of the measure of central tendency and the prior sample size, with $\theta_0^M = \theta_0^E$ and $n_0^M = n_0^E$.

5. A Shiny/R App to Elicit the Hyperparameters

The elicitation procedures described above can be used to formalize the belief that

it is highly plausible that θ belongs to a specific interval. To provide a user-friendly tool to easily implement the procedures, we developed a Shiny/R app available at https://vales.shinyapps.io/betapriors_app/. The app allows to choose between the mode or the mean as prior measure of central tendency. Then, once a numeric value has been entered for this measure, the user can dynamically appreciate how the shape of the beta prior distribution changes according to the prior sample size value passed through a slider. Moreover, the lower and upper limits of an interval of interest can be provided and the prior probability assigned to the interval is highlighted and indicated in the plot of the prior density. Thus, the prior sample size can be selected according to the level of this prior probability we wish to achieve.

5.1. Example 1

Let us consider an example provided by Spiegelhalter *et al.* [13], where the response rate of an experimental drug is supposed to lie between 0.2 and 0.6 on the basis of previous experience. The Authors translated this information into the beta prior of hyperparameters 9.2 and 13.8, by exploiting a normal approximation for the beta distribution of mean 0.4 and imposing a 95% of prior probability assigned to [0.2, 0.6]. As an alternative, we can exploit the Shiny/R App to set the hyperparameters as in (3) by fixing the prior mean equal to 0.4 and selecting the prior sample size n_0^E through the slider in order to have a prior probability assigned to the interval of interest equal to 0.95 (see Figure 2). We obtain the prior density $\pi(\theta) = Beta(\theta; 8.6, 12.9)$ with hyperparameters slightly different from the ones obtained with the procedure based on the use of a normal approximation, that actually provides a prior probability that $\theta \in [0.2, 0.6]$ equal to 0.958.

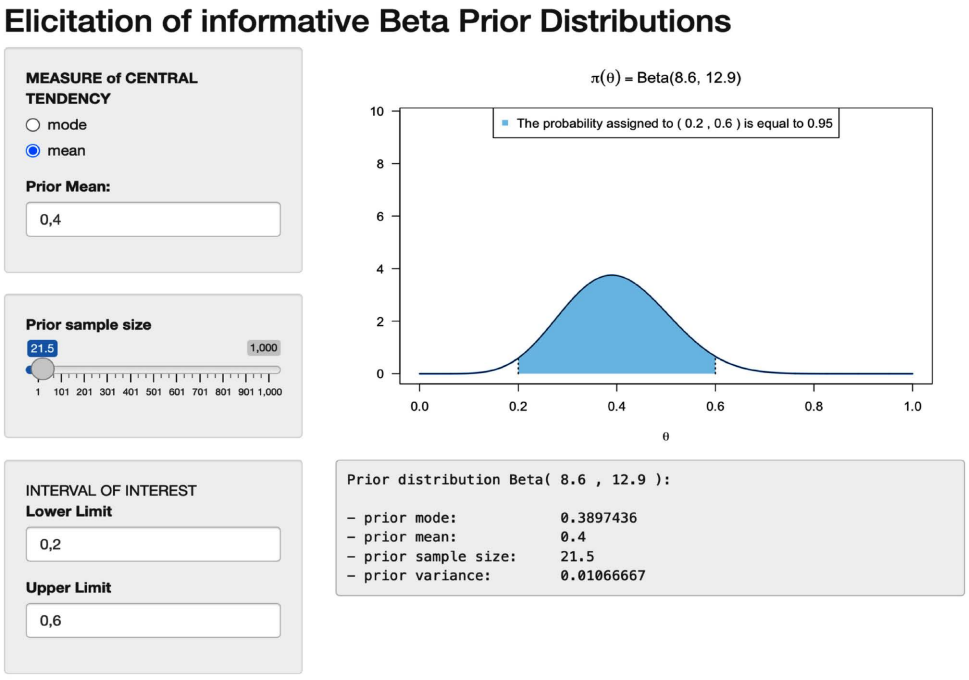

Figure 2. Shiny app user interface to elicit a beta prior distribution with mean equal to 0.4 that assigns a 95% prior probability to the interval [0.2, 0.6].

5.2. Example 2

Let us assume that our prior knowledge suggests that the success probability θ cannot take certain values or that we need to elicit a beta prior with support smaller than the parameter space [0, 1]. The latter case occurs, for instance, at the design stage of a one-sample experiment when a hybrid frequentist-Bayesian approach is used to determine the appropriate sample size. When the focus is on power analysis, the frequentist criterion selects the minimal sample size that guarantees a given power, for a fixed Type I error rate, conditional on the assumption that the true θ is equal to a clinically relevant *design value*, θ^D, that belongs to the alternative hypothesis. To overcome local optimality, implied by the fixed design value, the hybrid approach assigns a prior distribution to θ, typically called *design prior distribution*, to realize the assumption that the alternative hypothesis is true [14]. The sample size criteria is then based on the average of the traditional power over the design prior. For an exhaustive discussion about this topic the readers are referred to [15].

Specifically, let us focus on the hypotheses $H_0 : \theta \le 0.3$ and $H_1 : \theta > 0.3$. In order to apply the hybrid approach to determine the optimal sample size, we need to elicit a beta design prior for θ that 1) assigns negligible probability to values of θ outside the interval $(0.3,1]$ and 2) has mode equal to the design value that we would have set if we had used the classical approach. The Shiny/R app can be used to easily obtain such a prior density. Assuming that $\theta^D = 0.4$, a first way of proceeding consists in setting the prior mode equal to 0.4 and specifying the limits of the interval of interest equal to 0.3 and 1, respectively. Then, we use the

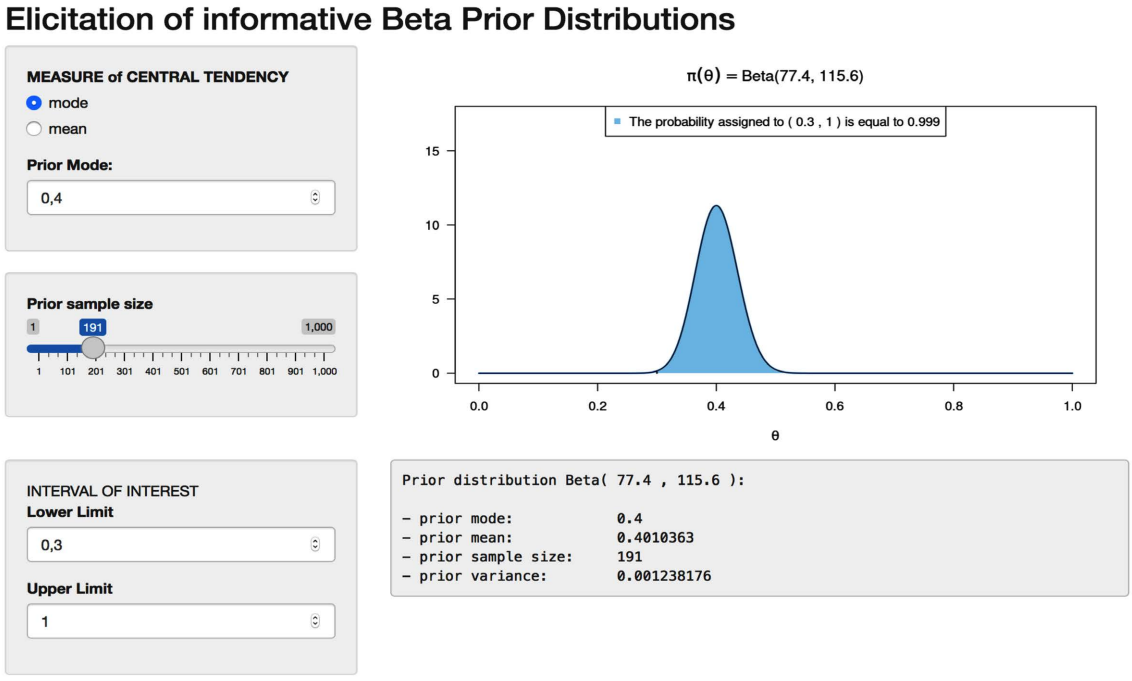

Figure 3. Shiny app user interface to elicit a beta prior distribution with mode equal to 0.4 that assigns a 999% prior probabability to the interval [0.3, 1].

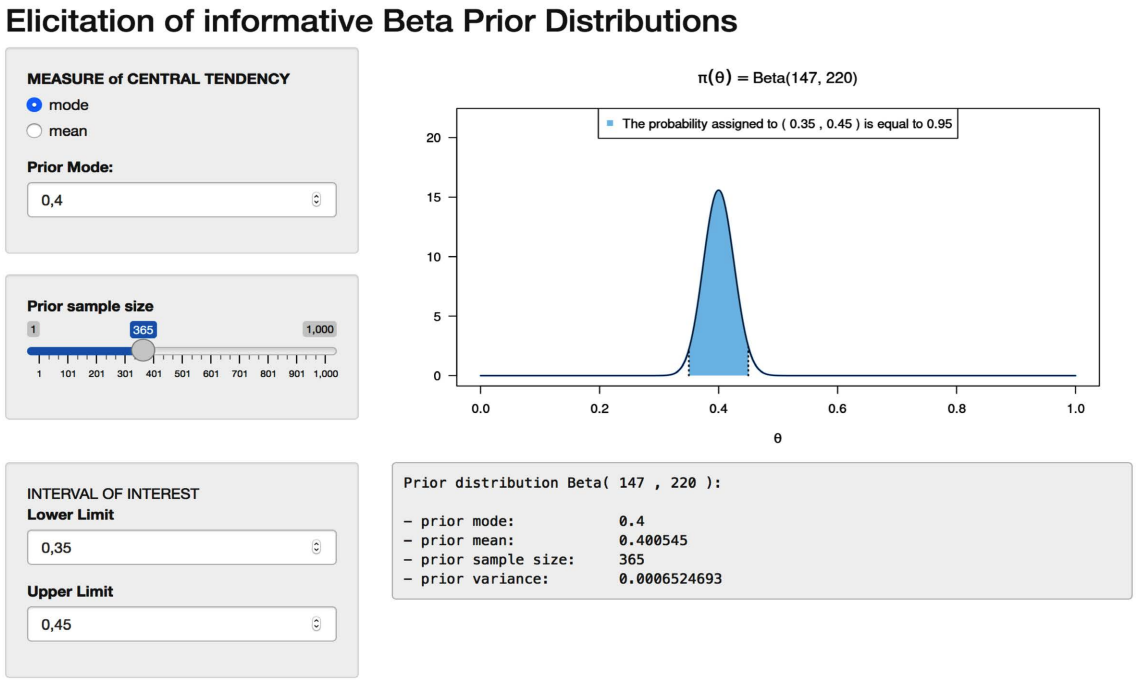

Figure 4. Shiny app user interface to elicit a beta prior distribution with mode equal to 0.4 that assigns a 95% prior probability to the interval [0.35, 0.45].

slider to increase the prior sample size, n_0^M, until the prior probability assigned to $(0.3,1]$ achieves the level 0.999. **Figure 3** shows that the prior obtained is $\pi(\theta) = Beta(\theta; 77.4, 115.6)$, based on a prior sample size equal to 191. As an alternative, we can choose as interval of interest an interval centred at θ^D and entirely belonging to the alternative hypothesis. In this case, the prior sample size can be fixed as the smallest value such that the prior probability of the interval reaches a desired high value, *i.e.* 0.95. For instance, by using the shiny app, we obtain that the beta prior $\pi(\theta) = Beta(\theta; 147, 220)$ has mode in 0.4 and assigns a prior probability equal to 0.95 to the interval $[0.35, 0.45]$, being based on a prior sample size equal to 365 (see **Figure 4**). This latter prior is less dispersed around θ^D, being based on a larger prior sample size, and thus expresses less uncertainty on the guessed design value.

6. Conclusions

Binary outcomes are dichotomous data that can take exactly two possible values and are often analysed in many fields. In the Introduction, we specifically refer to epidemiology and clinical trials, but many applications concern different context, such as, marketing, education, environment and social science. The statistical analysis of these data is often conducted by exploiting the Bayesian approach that is very popular and widely used nowadays. Many introductory textbooks and papers on Bayesian procedures cover binary data analysis, because it is really simple, but they typically provide only one of the two possible interpretations of the hyper-parameters of the beta prior distribution as extra data. This paper clarifies

why both the interpretations are reasonable and, therefore, it is particularly aimed at non-expert users of Bayesian statistics and for students who might be confused by inconsistent definitions across sources. By explicitly discussing both interpretations and their implications, we aim to provide a clearer and more coherent understanding of the beta prior density, supporting more informed and consistent applications in practice.

Moreover, the great popularity of Bayesian methods is in part due to the proliferation of user-friendly software tools available through open source platforms. We developed a very intuitive Shiny/R app that allows practitioners unfamiliar with Bayesian methods to easily apply the described elicitation procedures. Although the Shiny app itself is not the central contribution of this work, we think it can be useful as a support tool to operationalize the theoretical insights discussed. The app complements the methodological content by offering a dynamic and interactive means to explore the effects of different elicitation choices. In particular, it aims at fostering a better understanding of how prior parameters influence the beta prior shape and at clarifying the role of the prior sample size.

Note that the elicitation procedures based on the two interpretations lead to very similar beta prior densities for high values of the prior sample sizes and to very similar posterior analysis for large samples. Instead, in small-sample contexts, the posterior distribution may be noticeably sensitive to the prior sample size, and hence to whether it is defined as $\alpha+\beta-2$ (mode-based) or $\alpha+\beta$ (mean-based). Furthermore, as a limitation of the elicitation procedures presented, we can mention the fact that they assume a unimodal and well-defined belief about the success probability. These methods may not be suitable when the prior knowledge is vague, multi-modal, or intentionally uniform over a restricted range. In such cases, alternative elicitation strategies or prior structures (e.g., mixtures of beta distributions or truncated distributions) should be considered.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Armitage, P., Berry, G. and Matthews, J.N.S. (2002) Statistical Methods in Medical Research. 4th Edition, Wiley-Blackwell. https://doi.org/10.1002/9780470773666

[2] Christensen, R., Johnson, W., Branscum, A. and Hanson T. E. (2010) Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians. CRC Press.

[3] Rosner, G.L., Laud, P.W. and Johnson, W.O. (2021) Bayesian Thinking in Biostatistics. Chapman & Hall/CRC.

[4] Longford, N.T. (2021) Statistics for Making Decisions. Chapman & Hall/CRC Press.

[5] Lesaffre, E. and Lawson, A.B. (2012) Bayesian Biostatistics. John Wiley & Sons.

[6] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin D.B. (2013) Bayesian Data Analysis. Chapman & Hall/CRC.

[7] Blitzstein, J.K. and Hwang, J. (2019) Introduction to Probability. 2nd Edition, Chap-

man & Hall/CRC.

[8] Fox, J.P. (2010) Bayesian Item Response Modeling: Theory and Applications. Springer.

[9] Sambucini, V. (2008) A Bayesian Predictive Two-stage Design for Phase II Clinical Trials. *Statistics in Medicine*, **27**, 1199-1224. https://doi.org/10.1002/sim.3021

[10] Matano, F. and Sambucini, V. (2016) Accounting for Uncertainty in the Historical Response Rate of the Standard Treatment in Single-Arm Two-Stage Designs Based on Bayesian Power Functions. *Pharmaceutical Statistics*, **15**, 517-530. https://doi.org/10.1002/pst.1788

[11] Tan, S. and Machin, D. (2002) Bayesian Two-Stage Designs for Phase II Clinical Trials. *Statistics in Medicine*, **21**, 1991-2012. https://doi.org/10.1002/sim.1176

[12] Teramukai, S., Daimon, T. and Zohar, S. (2012) A Bayesian Predictive Sample Size Selection Design for Single-Arm Exploratory Clinical Trials. *Statistics in Medicine*, **31**, 4243-4254. https://doi.org/10.1002/sim.5505

[13] Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P. (2003) Bayesian Approaches to Clinical Trials and Health-care Evaluation. Wiley. https://doi.org/10.1002/0470092602

[14] Sambucini, V. (2017) Bayesian vs Frequentist Power Functions to Determine the Optimal Sample Size: Testing One Sample Binomial Proportion Using Exact Methods. In: Tejedor, J.P., Ed., *Bayesian Inference*, InTech, 77-97. https://doi.org/10.5772/intechopen.70168

[15] Kunzmann, K., Grayling, M.J., Lee, K.M., Robertson, D.S., Rufibach, K. and Wason, J.M.S. (2021) A Review of Bayesian Perspectives on Sample Size Derivation for Confirmatory Trials. *The American Statistician*, **75**, 424-432. https://doi.org/10.1080/00031305.2021.1901782