Method to Obtain a Vector of Hyperparameters: Application in Bernoulli Trials

Método para obtener un vector de hiperparámetros: aplicación en ensayos Bernoulli

Llerzy Torres[1,a], José Rafael Tovar Cuevas[1,b]

[1]School of Statistics, Universidad del Valle, Cali, Colombia

---

Abstract

The main difficulties when using the Bayesian approach are obtaining information from the specialist and obtaining hyperparameters values of the assumed probability distribution as representative of knowledge external to the data. In addition to the fact that a large part of the literature on this subject is characterized by considering prior conjugated distributions for the parameter of interest. An method is proposed to find the hyperparameters of a nonconjugated prior distribution. The following scenarios were considered for Bernoulli trials: four prior distributions (Beta, Kumaraswamy, Truncated Gamma and Truncated Weibull) and four scenarios for the generating process. Two necessary, but not sufficient conditions were identified to ensure the existence of a vector of values for the hyperparameter. The Truncated Weibull prior distribution performed the worst. The methodology was used to estimate the prevalence of two transmitted sexually infections in an Colombian indigenous community.

***Key words*:** Laplace's method; Bayesian inference; System of nonlinear equations.

Resumen

Las principales dificultades cuando se utiliza el enfoque Bayesiano son la obtención de información del especialista y la obtención de valores de los hiperparámetros de la distribución de probabilidad asumida como representante del conocimiento a priori. Adicionalmente, gran parte de la literatura sobre este tema considera distribuciones a priori conjugadas para el parámetro de interés. Un método es propuesto para encontrar los valores de los hiperparámetros de una distribución a priori no conjugada. Los siguientes escenarios son considerados para ensayos Bernoulli: cuatro distribuciones

[a]Posgraduate student. E-mail: llerzy.torres@correounivalle.edu.co

[b]Ph.D. E-mail: jose.r.tovar@correounivalle.edu.co

a priori (Beta, Kumaraswamy, Gamma Truncada y Weibull Truncada) y cuatro escenarios para el proceso generador. Dos condiciones necesarias, pero no suficientes fueron identificadas para asegurar la existencia de un vector de valores para los hiperparámetros. La distribución a priori Weibull Truncada fue la que peor desempeño presentó. La metodología fue utilizada para estimar la prevalencia de dos infecciones de transmisión sexual en una comunidad indígena de Colombia.

Palabras clave: Método de Laplace; Inferencia bayesiana; Sistema de ecuaciones no lineales.

1. Introduction

When Bayesian inference methods are used for obtaining information about a phenomenon of interest, the elicitation of the prior probability distribution is a key process. According to Garthwaite, Kadane & O'Hagan (2005), the statistician or facilitator establishes contact with an individual considered to be a specialist (expert) on the research topic. The expert has valuable subjective information about the data-generating process, having acquired experience from repeated exposure to the phenomenon. In order to incorporate the information about the data-generating process provided by the expert in the statistical model proposed for the study, the statistician has to "express" such information as a probability distribution (prior distribution) to establish "communication" with the information contained in the data, which is expressed as a likelihood function.

Expressing the prior information as a probability distribution implies two challenges for the facilitator: to establish a theoretical model of probability that best represents the natural behavior of the indicator assumed to be an expression of the data-generating process conditions in the statistical model and to determine the numerical values of the parameters associated with the theoretical model of probability (hyperparameters), to ensure that this model is an adequate representation of the conditions under which the data-generating process occurs (phenomenon of interest).

Different authors have studied the conditions considered in an elicitation process. Some authors such as Tversky & Kahneman (1973), Tversky & Kahneman (1974), Hogarth (1975) and Hogarth (1987), have studied the cognitive and psychological processes present in the individual who performs the experts functions when responding to interrogations by the facilitator related to the generating process. Winkler (1967) and Fowler & Floyd (1995) have studied the problem of how to formulate questions and what to ask the expert concerning the given issue. From a more probabilistic and statistical point of view, Murphy & Winkler (1974), Winkler (1967) and Kadane & Winkler (1988) have developed mathematical methods for obtaining different types of probability distributions that serve as representations of the level of presence-absence of prior information in the generating process. Likewise, Chaloner & Duncan (1983), Sindhu, Feroze & Aslam (2013), Tovar (2012), Penha (2014) and Flórez & Correa (2015) have worked on the development of mathematical procedures for obtaining the hyperparameter

values that best represent the conditions of the data-generating process in the prior theoretical distribution model. On the other hand Moala & Penha (2016), have proposed a simulation methodology for evaluating the effectiveness of these procedures.

One of the most used method to obtain the vector of hyperparameters ϕ of prior distribution π, is to evaluate a quantiles interval from the specialist and calculate from this its midpoint α and a fraction β of its rank. Using this, information and the moments of the prior distribution π, the system of equations is constructed

$$\begin{cases} E[\theta \mid \phi] &= \alpha \\ Var[\theta \mid \phi] &= \beta \end{cases}$$

The vector ϕ^* that satisfies the system defines a single prior distribution $\pi(\theta \mid \phi^*)$ as representative of the beliefs of the specialist. Given that the assumed values for the mean and variance of the subjective distribution are fixed, this method ends up forcing the specialist to accept the prior distribution obtained or provide a new quantiles interval that generates a different vector ϕ^*.

To avoid these difficulties, some authors have used Prior Predictive Distribution (PPD) to obtain the hyperparameter vector. Chaloner & Duncan (1983), considers Bernoulli trials, the prior Beta distribution, a quantile interval evaluated by the specialist and the PPD. On the other hand, survival times modeled with the Weibull distribution are used by Penha (2014), who assumes the bivariate Gamma distribution as a prior (nonconjugated) and constructs a system of equations with PPD approximations.

Except for Penha (2014), the methods of elicitation proposed by the aforementioned authors are characterized by assuming the conjugate of the likelihood function as prior distribution and obtaining information about the moments or percentiles of said distribution from the specialist to form a system of equations that provides the value of the hyperparameters. In this article this subject is addressed and a method to obtain hyperparameter values of nonconjugated prior distributions is proposed. Using the Laplace (1773) method, we developed asymptotic approximations of the PPD obtaining a system of nonlinear equations, which is solved using some numerical approximation as the Newton-Raphson method. The Laplace Method (LM), which has been studied extensively by Tierney & Kadane (1986), Azevedo-Filho & Shachter (1994) and Vidal (2014) in order to approximate posterior moments, marginal distributions and the Bayes factor, among others. In our literature review, we found one work that implemented the LM to obtain values of the hyperparameters of a prior distribution, (Penha 2014).

To illustrate the proposed method, a realization of a random sample of independent variables with a Bernoulli distribution is considered and four priors (Beta, Kumaraswamy, Truncated Gamma, Truncated Weibull) were assumed. A simulation study using four different values of the parameter to be estimated was carried out and the statistical performance of the Bayesian estimates was compared with those obtained using the classical approach. Data obtained from a study

carried to estimate the prevalence of HIV and Syphilis infection in a Colombian indigenous community, were used to illustrate the performance of the proposed method.

2. Methodological Proposal to Obtaining Hyperparameter Values

In general the Figure 1 presents the steps of the method described in this section.
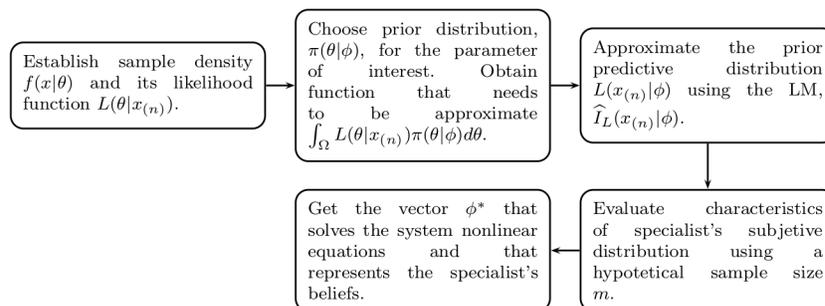

FIGURE 1: Short representation of the proposed procedure.

The procedure used to approximate integrals is that developed by Laplace (1773) and formalized by authors such as Erdélyi (1956) and Bruijn (1961). The Laplace method formally indicates: Let $h(.)$ and $b(.)$ be analytical and smooth functions infinitely differentiable over a domain Ω, so that their Taylor series expansion around the local minimum $\widehat{\theta}$ of the convex function $h(.)$ in Ω exists. Then, for a considerably high value of n, the integral

$$I_n = \int_\Omega b(\theta) \exp\{-nh(\theta)\}d\theta$$

has the following result

$$\widehat{I}_n = (2\pi)^{1/2} n^{-1/2} b(\widehat{\theta}) \mid \Sigma \mid^{1/2} \exp\{-nh(\widehat{\theta})\}\{1 + O(n^{-1})\}$$

where, $\Sigma = \left(\dfrac{\partial^2 h(\widehat{\theta})}{\partial \theta^2} \right)^{-1}$ and $O(n^{-1})$ is the Landau notation and indicates that the following is fulfilled for a function $g(n) = O(n^{-1})$; namely,

$$(\exists n_0 \in N)(\forall n \geq n_0)(\exists \epsilon > 0)(\mid g(n) \mid \leq \epsilon \mid n^{-1} \mid)$$

2.1. Obtaining the Hyperparameter Values

Consider that an experiment has been performed n times (in a heuristic way, with a computational simulation algorithm, through real realizations) so

$x_{(n)} = \{x_1, \ldots, x_n\}$ which is assumed at least exchangeable, i.e. each observation having distribution $f(x \mid \theta)$, where θ is an unknown scalar or vector of parameters that represents the conditions under which a generating process occurs. The information about θ contained in the sample of observations can be incorporated into the model through the likelihood function $L(\theta \mid x_{(n)})$, which has the following form,

$$L(\theta \mid x_{(n)}) = \prod_{i=1}^{n} f(x_i \mid \theta) \tag{1}$$

In addition to the sample information, it is possible to consider in the construction of the statistical model that represents the situation under study, information "external" to the sample that has been obtained from the predictions of Expert A in the area of research and from the parametric space Ω that contains all the values of θ. Thus a probability distribution function, $\pi(\theta \mid \phi)$ is sought, where ϕ is a hyperparameter vector of l components, whose values will have to be found to represent A's predictions. On some occasions, the Ω domain can be restrictive in selecting the prior; but it can be ignored by using truncated distributions or re-parameterizations.

The information represented in the likelihood function can be combined with the expression through $\pi(\theta \mid \phi)$, using Bayes' formula, thereby obtaining the posterior distribution (2).

$$\pi(\theta \mid x_{(n)}) = \frac{L(\theta \mid x_{(n)})\pi(\theta \mid \phi)}{\int_{\Omega} L(\theta \mid x_{(n)})\pi(\theta \mid \phi)d\theta} = \frac{L(\theta \mid x_{(n)})\pi(\theta \mid \phi)}{L(x_{(n)} \mid \phi)} \tag{2}$$

The denominator of (2) represents the marginal likelihood of (1), also called the PPD. Finding an analytical expression for (2) depends on the prior distribution used. To avoid this constraint, the asymptotic approximation provided by LM was used. The asymptotic approximation of the denominator in (2) is denoted as $\widehat{I}_L(x_{(n)} \mid \phi)$, where n represents the sampling size and allows determining the approximation accuracy; while ϕ is the constant vector that represents the hyperparameters of the prior distribution.

To implement the LM, the functions $b(\theta)$ and $h(\theta)$, which represent the PPD, must be chosen; thus $b(\theta)\exp\{-nh(\theta)\} = L(\theta \mid x_{(n)})\pi(\theta \mid \phi)$. According to the literature, two parameterizations can be considered (selections of b and h). The first parameterization (3) is based on classical techniques, given that maximizing $-nh(\theta)$ is equivalent to maximizing the logarithm of the likelihood function. However, Kass & Raftery (1995) indicated that the approximation error increases when the prior distribution is informative.

$$-nh(\theta) = \log L(\theta \mid x_{(n)}); \qquad b(\theta) = \pi(\theta \mid \phi) \tag{3}$$

On the other hand, in the second parameterization (4), the maximum value of $-nh(\theta)$ is called the posterior mode because it is the result of maximizing the logarithm of the product between the likelihood and the prior distribution; that is, the logarithm of the non-normalized posterior distribution. The latter parameterization is the one implemented in the proposed method.

$$-nh(\theta) = \log L(\theta \mid x_{(n)}) + \log \pi(\theta \mid \phi); \qquad b(\theta) = 1 \tag{4}$$

Once the prior distribution has been chosen and $\widehat{I}_L(x_{(n)} \mid \phi)$ has been obtained, the information from Expert "A" is extracted as follows:

i) A value m is selected for the number of elements that a hypothetical sample contains.

ii) Using m, characteristics of the expert's subjective distribution are evaluated, as: quantiles $x_{(m),j}$, probabilities α_j, correlation or association coefficient ρ_j, mean, among other descriptive measures. The number of characteristics that must be evaluated coincides with the number l of components of ϕ.

iii) Trial samples are generated that satisfy each evaluated characteristic and the following system of l equations with l unknowns (components of ϕ) is constructed.

$$\widehat{I}_L(x_{(m),1} \mid \phi) - \alpha_1 = 0$$
$$\widehat{I}_L(x_{(m),2} \mid \phi) - \alpha_2 = 0$$
$$\vdots$$
$$\widehat{I}_L(x_{(m),l} \mid \phi) - \alpha_l = 0$$

For example, consider a sample size (hypothetical) m and a vector of hyperparameters of $l = 2$ components. Two quantiles have been evaluated from the expert, $x_{(m),1}$ and $x_{(m),2}$, with the probabilities α_1 and α_2, respectively. Using this information, the specialist's Prior Predictive Pseudo-Distribution (PPPD) is defined, as:

$$L(x_{(m),j} \mid \phi = (a,b)) = \int_\Omega L(\theta \mid x_{(m),j})\pi(\theta \mid \phi)d\theta$$

where $L(\theta \mid x_{(m),j})$ is the likelihood function defined by the trial sample of size m containing the quantile $x_{(m),j}$. With this fact, the PPPD depends on the vector of hyperparameters ϕ. Thus, the method of obtaining hyperparameters consists of finding a vector of constants ϕ^* that satisfies the following equation:

$$\widehat{I}_L(x_{(m),j} \mid \phi^*) = \alpha_j \qquad \text{for } j = 1, 2$$

The vector ϕ^* representing the beliefs of the expert is obtained as the solution of system (5). Several methods may be used to determine such solutions, such as successive approximations, Newton-Raphson and variants (such as finite differences, Jacobi, Gauss-Sediel, and Quasi-Newton, among others). The Newton-Raphson method will be used as it exhibits quadratic convergence and other properties that were studied in detail by Dennis-Jr & Schnabel (1996).

$$\begin{aligned}\widehat{I}_L(x_{(m),1} \mid \phi) - \alpha_1 = 0 \\ \widehat{I}_L(x_{(m),2} \mid \phi) - \alpha_2 = 0\end{aligned} \tag{5}$$

3. n Bernoulli Observations

Let $y_1, \ldots, y_n$ be a sample of n observations obtained when performing n independent Bernoulli experiments, each with a $\theta \in [0, 1]$ probability of success (the sample can be only exchangeable). The sufficient statistic $x = \sum_{i=1}^{n} y_i$ is defined. Given the characteristics of the experiment, it is known that $X \sim Binomial(n, \theta)$; thus the likelihood function of the parameter θ related to the sufficient statistic x will be

$$L(\theta \mid x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \tag{6}$$

Thus for any prior distribution $\pi(\theta \mid \phi)$, the PPD will be

$$L(x \mid \phi) = \int_\Omega \binom{n}{x} \theta^x (1-\theta)^{(n-x)} \pi(\theta \mid \phi) d\theta \tag{7}$$

To approximate $L(x \mid \phi)$, the LM is used, which requires the selection of functions $-nh(.)$ and $b(.)$, the calculation of the minimum value of $h(.)$ and the second derivative of $h(.)$. For the Bernoulli distribution, the parameterization (4) generates

$$-nh(\theta) = \log \binom{n}{x} + x \log \theta + (n-x) \log(1-\theta) + \log \pi(\theta \mid \phi) \tag{8}$$

If parameterization (3) is used, expression (8) would be obtained without the term $\log \pi(\theta \mid \phi)$ –that is, only the logarithm of the binomial density–; thus the maximum of $-nh(.)$ coincides with the maximum likelihood estimator, $\widehat{\theta} = x/n$. Determining an analytical expression for the maximum or posterior mode of (8) $\widehat{\theta}$ depends on the shape of the selected prior distribution. In some cases, it is impossible to obtain it. Therefore, it is necessary to apply the numerical approach method called Bisection to find the roots of the following expression:

$$-nh'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} + \frac{\pi'(\theta \mid \phi)}{\pi(\theta \mid \phi)} \tag{9}$$

Likewise, the asymptotic approximation of (7) by LM depends on n, $x = x_{(n)}$ and ϕ: Its analytical expression is:

$$\widehat{I}_L(x_{(n)} \mid \phi) \approx$$

$$(2\pi)^{1/2} \left| \underbrace{\frac{x}{\widehat{\theta}^2} + \frac{n-x}{(1-\widehat{\theta})^2} - \frac{\pi''(\widehat{\theta} \mid \phi)\pi(\widehat{\theta} \mid \phi) - [\pi'(\widehat{\theta} \mid \phi)]^2}{[\pi(\widehat{\theta} \mid \phi)]^2}}_{h''(\widehat{\theta})} \right|^{-1/2} \underbrace{\binom{n}{x} \widehat{\theta}^x (1-\widehat{\theta})^{(n-x)} \pi(\widehat{\theta} \mid \phi)}_{\exp\{-nh(\widehat{\theta})\}} \tag{10}$$

Four families of probability distribution are considered to model the prior information about the natural behavior of the parameter of interest. The first two

(Beta, Kumaraswamy) are characterized by modelling the natural behavior of the random variables whose domain is the interval $\Omega = [0, 1]$. The Beta distribution is the conjugate prior when a Binomial model is assumed for the likelihood function and its density function (11) is characterized by two form parameters, $a, b > 0$. If θ is a random amount of interest such that $\theta \in \Omega$, it can be assumed that $\theta \sim Beta(a, b)$ so that $\phi = (a, b)$, then:

$$\pi(\theta \mid \phi) = \frac{1}{Be(a, b)} \theta^{a-1}(1-\theta)^{b-1} \qquad \begin{array}{c} \theta \in [0, 1] \\ a, b > 0 \end{array} \tag{11}$$

When using (11) in (7), the PPD is obtained:

$$L(x \mid \phi) = \underbrace{\binom{n}{x} \frac{1}{Be(a, b)}}_{\Delta_{Bet}} \int_{\Omega} \theta^{x+a-1}(1-\theta)^{(n+b-x-1)} d\theta \tag{12}$$

The integrand of (12) may be a Beta distribution, and the closed form (13) can be obtained.

$$L(x \mid a, b) = \Delta_{Bet} Be(x + a, b + n - x) \tag{13}$$

Other prior distribution assumed is the Kumaraswamy (Kum), with density function $\pi(\theta \mid a, b) = ab\theta^{a-1}(1-\theta^a)^{b-1}$, with $a, b > 0$ form parameters. It has the advantage of having a probability distribution that is analytically more treatable than Beta. This distribution is applicable to model phenomena, whose results have lower and higher limits, such as the weight of individuals, scores obtained from tests (Sindhu et al. 2013).

Besides of Kum and Beta distributions, we used as prior distributions, the truncated Gamma and the truncated Weibull. Given that, both correspond to theoretical models constructed for random variables that take non-negative values, for effects of the elicitation method when using Bernoulli trials, each one was truncated in the interval $\Omega = [0, 1]$ considering the process associated to probability of success $\theta \in \Omega$. The Gamma distribution, is characterized by having two parameters: one of form $a > 0$ and the other of inverse scale $b > 0$, also called a rate parameter or event rate. The expression of the Truncated Gamma (TG) distribution (14) contains a normalization constant (K), which depends of the values of the hyperparameters $\phi = (a, b)$.

$$Tr(\theta \mid a, b) = K^{-1}(a, b)\theta^{a-1}e^{-b\theta}; \qquad K(a, b) = \int_0^1 \theta^{a-1}e^{-b\theta}d\theta \tag{14}$$

The last prior probability distribution considered is the Weibull, which has two parameters: one of form $a > 0$ and the other of scale $b > 0$. The Truncated Weibull (TW) distribution, involves a normalization constant (K) that depends of the hyperparameters $\phi = (a, b)$.

$$Tr(\theta \mid a, b) = K^{-1}(a, b)\theta^{a-1}e^{-(\theta/b)^a}; \qquad K(a, b) = \int_0^1 \theta^{a-1}e^{-(\theta/b)^a}d\theta \tag{15}$$

The last two distributions were considered under the assumption that specialist thinks the probability of success takes values closed to zero in most of times.

With each of these prior distributions, the LM is implemented to determine an analytical approximation of the PPD (10). These expressions are presented in Table 2 and generate a surface of prior predictive probabilities that can be plotted on a region $D \subseteq R^2$. For example, consider 10 successes obtained in 50 independent repetitions of a Bernoulli experiment, with $D = [0, 100]^2$ and the prior Beta as representative of a specialist's beliefs. On the left side of Figure 2, the approximation of the PPD (10) appears in blue and its closed form in red (13); while on the right side the approximation of the PPD (10) is presented, assuming the prior Kumaraswamy. These approximations are implemented to form the system of nonlinear equations (5).
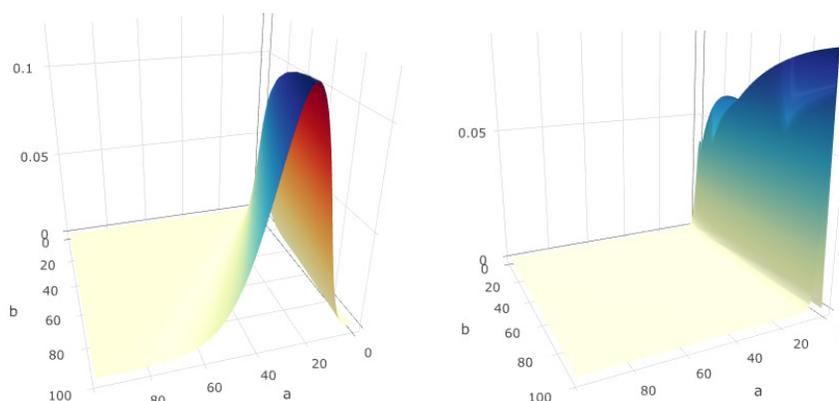

FIGURE 2: Surface of the PPD for 10 successes in 50 repetitions of independent Bernoulli experiments with Beta (left) and Kumaraswamy (right) prior distribution.

To obtain information through an individual who is an expert in the matter of interest, the following sentences can be posed:

a) Suppose that there is a hypothetical sample of m elements. Indicate two values: a minimum and a maximum $(x_{(m),1}, x_{(m),2})$ that represent the number of elements with the characteristic of interest. *(Example: For a sample of size $m = 50$, the expert has indicated that it would be expected to obtain a minimum of $x_{(m),1} = 5$ and a maximum of $x_{(m),2} = 10$ elements, respectively with the characteristic of interest)*

b) Now consider that there are 100 groups, each one with m elements. In how many groups would you expect to have exactly $x_{(m),1}$ and $x_{(m),2}$ elements with the characteristic of interest, respectively? *(Example: Each of the 100 groups has 50 elements. The expert expects to have 8 groups with $x_{(m),1} = 5$ elements that present the characteristic of interest, resulting in a probability of occurrence $\alpha_1 = 8/100$ for $x_{(m),1}$. Likewise, $\alpha_2 = 9/100$ indicates that the expert would expect to have $x_{(m),2} = 10$ elements with the characteristic of interest in 9 of the 100 groups)*

c) What is the number of elements $x_{(m),mod}$ that you consider would present the characteristic of interest with more frequency? This last question is related to the mode value in order to verify the consistency of the experts responses; therefore, $x_{(n),mod}$ has to remain in the interval $[x_{(m),j}, x_{(m),j'}]$ for $j < j'$. *(Example: The expert indicates that the number of elements with the characteristic of interest that will occur with higher frequency is $x_{(m),mod} = 8$. Then, with the information obtained from the previous items a) and b), this is verified that $8 \in [5, 10]$. Thus the expert's information is coherent)*

Using the data presented in items (a) and (b), the system of nonlinear equations is defined (5), dependent on the vector $\phi = (a, b)$.

$$\begin{aligned}
\widehat{I}_L(x_{(50),1} = 5 \mid \phi = (a, b)) - 8/100 = 0 \\
\widehat{I}_L(x_{(50),2} = 10 \mid \phi = (a, b)) - 9/100 = 0
\end{aligned} \tag{16}$$

The expression $\widehat{I}_L(x_{(50),j} \mid \phi)$ represents the PPPD given by (10) for prior distribution and a pair of hyperparameter values ϕ. To determine the solution vector ϕ^*, the Newton-Raphson method is used, and the initial point is obtained graphically through the system-level curves. Table 1 provides the most relevant results, such as the number of iterations, error vector and solution vector (value of hyperparameters). The results of the prior TW distribution are not presented because the PPPD was lower than the one provided by the specialist.

TABLE 1: Summary of the Newton-Raphson method for hyperparameters obtained considering three prior distributions.

Prior	Initial Point	N° Iter.	$(\widehat{I}_L(x_{(m),1} \mid \phi^*) - \alpha_1,$ $\widehat{I}_L(x_{(m),2} \mid \phi^*) - \alpha_2)$	Vector Solution (ϕ^*)
Beta	$(8, 36)$	5	$(-3.9 \times 10^{-10}, -2.2 \times 10^{-9})$	$(8.26, 39.08)$
Kumaraswamy	$(3, 90)$	14	$(-1.2 \times 10^{-10}, -8.1 \times 10^{-13})$	$(3.35, 257.98)$
TG	$(14, 78)$	7	$(-5.3 \times 10^{-11}, 3.5 \times 10^{-11})$	$(10.39, 59.39)$

In the process of obtaining the vector solution ϕ^* of the system of nonlinear equations (16), two necessary conditions were identified but were not sufficient to guarantee the existence of the solution. The first indicates that $\widehat{I}_L(x_{(m),j} \mid \phi)$ must not be exceeded by the specialist's probabilities α_j, $j = 1, 2$; and the second is that the amplitude of the interval evaluated $(x_{(m),1}, x_{(m),2})$ should not be too width.

In this table the following expressions are used:

$$k_1 = \frac{c_1}{\theta} - \frac{c_2}{1 - \theta} \qquad k_2 = -\frac{c_1}{\theta^2} - \frac{c_2}{(1 - \theta)^2} \qquad k_3 = c_3 a \theta^{a-2} \frac{(a - 1)(1 - \theta^a) + a\theta^a}{(1 - \theta^a)^2}$$

with $c_1 = x + a - 1$, $c_2 = n - x$, $c_3 = b - 1$ y $\phi = (a, b)$.

TABLE 2: Approximation of prior predictive distributions.

Prior	$L(x \mid \phi)$		$\widehat{I}_L(x \mid \phi)$
Beta	$\binom{n}{x}\dfrac{1}{Be(a,b)}\,\Delta_{Bet}\displaystyle\int_\Omega \theta^{c_1}(1-\theta)^{c_2}d\theta \quad \Delta_{Bet} =$	$-nh'(\theta)=k_1-\dfrac{b-1}{1-\theta}$ $-nh''(\theta)=k_2-\dfrac{\frac{b-1}{1-\theta}-1}{(1-\theta)^2}$	$(2\pi)^{1/2}\dfrac{\left|c_1 c_2\right|^{1/2}c_1^{c_1}c_2^{c_2}}{(c_1+c_2)^{3/2+c_1+c_2}}\Delta_{Bet}$
Kumaraswamy	$\Delta_{Kum}\displaystyle\int_\Omega \theta^{c_1}(1-\theta)^{c_2}(1-\theta^a)^{c_3}d\theta$ $\Delta_{Kum}=ab\binom{n}{x}$	$-nh'(\theta)=k_1-\dfrac{c_3 a\theta^{a-1}}{1-\theta^a}$ $-nh''(\theta)=k_2-k_3$	$\dfrac{\widehat{\theta}^{c_1}(1-\widehat{\theta})^{c_2}(1-\widehat{\theta}^a)^{c_3}\left|h''(\widehat{\theta})\right|^{-1/2}}{(2\pi n^{-1})^{-1/2}\Delta_{Kum}^{-1}}$
Truncated Gamma	$\Delta_{TG}\displaystyle\int_\Omega \theta^{c_1}(1-\theta)^{c_2}e^{-b\theta}d\theta \quad \Delta_{TG} =$ $K^{-1}(a,b)\binom{n}{x}$	$-nh'(\theta)=k_1-b$ $-nh''(\theta)=k_2$	$\dfrac{\widehat{\theta}^{c_1}(1-\widehat{\theta})^{c_2}e^{-b\widehat{\theta}}\left|h''(\widehat{\theta})\right|^{-1/2}}{(2\pi n^{-1})^{-1/2}\Delta_{GT}^{-1}}$
Truncated Weibull	$\Delta_{TW}\displaystyle\int_\Omega \theta^{c_1}(1-\theta)^{c_2}e^{-(\theta/b)^a}d\theta \quad \Delta_{TW} =$ $K^{-1}(a,b)\binom{n}{x}$	$-nh'(\theta)=k_1-\dfrac{a}{b}\left(\dfrac{\theta}{b}\right)^{a-1}$ $-nh''(\theta)=k_2-\dfrac{a(a-1)}{b^2}\left(\dfrac{\theta}{b}\right)^{a-2}$	$\dfrac{\widehat{\theta}^{c_1}(1-\widehat{\theta})^{c_2}e^{-(\widehat{\theta}/b)^a}\left|h''(\widehat{\theta})\right|^{-1/2}}{(2\pi n^{-1})^{-1/2}\Delta_{WT}^{-1}}$

Given that the hypothetical sample size (m) delivered to the expert can affect the process to obtain hyperparameters, four scenarios of the generating process expressed as probabilities of success were established, and three hypothetical sample values were assumed for each one. These data were presented for the specialist and in the Table 3 presents some of these results, as the maximum values generated by each prior distribution on the surface of the PPPD. For these cases, a mesh $D = [0, 100]^2$ was used initially. These surface limited the mesh (Table 4). For example, on the right side of Figure 2, the surface of the PPPD generated by the prior Kumaraswamy was presented in the scenario of $\theta = 0.15$ and mesh $D = [0, 100]^2$; however, the region where the surface is far away from the ab plane is $[0, 15] \times [0, 100]$. Other aspects that can be evidenced from Table 3 are:

TABLE 3: Scenario of the generating process and the three sizes of hypothetical samples presented to the specialist.

Scenario θ	Case	Specialist's Beliefs			$L(\theta_j \mid x_{(m),j})$	$\max\limits_{(a,b)\in D} \widehat{I}_L(x_{(m),j} \mid \phi)$			
		m	$x_{(m),j}$	θ_j		Beta	Kum	TG	TW
0.150	1	100	10	0.100	0.132	0.100	0.064	0.090	0.014
			20	0.200	0.099	0.074	0.048	0.066	0.013
	2	50	5	0.100	0.185	0.154	0.114	0.147	0.028
			10	0.200	0.140	0.118	0.086	0.110	0.025
	3	10	1	0.100	0.387	0.371	0.350	0.365	0.122
			2	0.200	0.302	0.292	0.269	0.284	0.118
0.500	4	100	45	0.450	0.080	0.064	0.041	0.048	0.012
			55	0.550	0.080	0.064	0.044	0.045	0.013
	5	50	23	0.460	0.113	0.100	0.072	0.081	0.023
			28	0.560	0.113	0.100	0.077	0.078	0.025
	6	10	4	0.400	0.251	0.245	0.219	0.232	0.102
			6	0.600	0.251	0.245	0.225	0.225	0.123
0.850	7	100	80	0.800	0.099	0.074	0.069	0.042	0.023
			90	0.900	0.132	0.100	0.100	0.049	0.042
	8	50	40	0.800	0.140	0.118	0.110	0.078	0.045
			45	0.900	0.185	0.153	0.158	0.093	0.079
	9	10	8	0.800	0.302	0.292	0.274	0.254	0.186
			9	0.900	0.387	0.371	0.354	0.331	0.283
0.900	10	100	85	0.850	0.111	0.082	0.082	0.043	0.030
			95	0.950	0.180	0.129	0.153	0.074	0.077
	11	50	43	0.860	0.161	0.135	0.133	0.082	0.060
			48	0.960	0.276	0.225	0.243	0.155	0.162
	12	10	8	0.800	0.302	0.292	0.274	0.254	0.186
			10	1.000	1.000	0.769	0.776	0.757	Inf

$L(\theta_j \mid x_{(m),j})$ is the function of likelihood that now depends of parameter θ_j and $x_{(m),j}$ is a j-th hypothetical sample of size m.

TABLE 4: Regions D for approximating the PPPD based on the LM.

Prior	0.150			0.500			0.850			0.900		
	1	2	3	4	5	6	7	8	9	10	11	12
Beta	$[0,100] \times [0,100]$											
Kum	$[0,15] \times [0,100]$			$[0,30] \times [0,100]$		$[0,100]^2$	$[0,100]^2$					
TG	$[0,100] \times [0,100]$											
TW	$[0,10] \times [0,100]$			$[0,25] \times [0,100]$		$[0,100]^2$	$[0,100]^2$					-

- The maximum PPPD values increase as the size of the hypothetical sample decreases so that the values obtained from this distribution generate an upper limit for the specialist's α_j.

- The information provided by the specialist generates a likelihood of beliefs, $L(\theta_j \mid x_{(m)j})$, which approximates the maximum PPPD values when the size of the hypothetical sample decreases.

- The prior distribution Beta and TW generated, respectively, the highest and lowest PPPD.

- When the parameter to be estimated had central values (that is, around 0.5) in its domain and the sum of the extremes of the interval evaluated by the specialist coincided with the size of the hypothetical sample considered $(x_{(m),1} + x_{(m),1} = m)$, the components of each of the following vectors were equal: the probability of the specialist $(\alpha_1 = \alpha_2)$ and maximum prior predictive probabilities, these last amounts appear by pairs of rows from the seventh to the tenth column of Table 3.

4. Simulation Study

Consider from the previous section the probabilities for item (b) fixed for each scenario of the parameter in Table 3. Figure 3 presents the form of the prior densities elicited and Table 5 the respective hyperparameters, descriptive measures and case from which the specialist's information was obtained. For example, in the scenario where θ takes a value in the middle of its domain ($\theta = 0.5$), based on the prior Beta, the case was considered where the specialist was evaluated based on a hypothetical sample of $m = 50$ (Case 5). With the Kumaraswamy and TG the hypothetical sample was $m = 10$ elements (Case 6); whereas with the TW, no solution to system 16 was found in any of the cases.

Assuming that the specialist agrees with the probability contained in the interval $[\theta_1, \theta_2]$, which was obtained from the prior distribution proposed, It was developed one simulation procedure. Bernoulli samples ($s = 1, \ldots, 10000$) were generated assuming different θ and sample size n_k (simulation's scenarios). In each of these scenarios, the probability of success was estimated. The mean, maximum, minimum and MSE (Mean Square Error) of the estimates were computed. The procedure was carried out 56 times, because four prior distributions, four values of

θ and four sample sizes $n_k = \{20, 50, 125, 312\}$ were considered. The sample sizes were associated with the first four elements of a geometric serie, whose common ratio was 2.5 and first term, 20. The algebraic expressions of the mean and posterior variance were obtained by approximating the $r-$th posterior moment using the Laplace Method. See Appendix A.

TABLE 5: Descriptive measures of prior distributions elicited for the scenarios of the generating process. Pr represents $P[\theta \in (\theta_1, \theta_2)]$.

| θ	Prior	ϕ^*		Measures of Central Tendency			Var	Pr	Case
		a	b	Mean	Median	Mode			
0.150	Beta	8.264	39.078	0.175	0.170	0.160	0.003	0.630	
	Kum	3.354	257.978	0.171	0.171	0.172	0.003	0.582	2
	TG	10.392	59.388	0.175	0.169	0.158	0.003	0.645	
	TW	1.225	3.212	0.531	0.540	NaN	0.077	0.088	3
0.500	Beta	149.993	140.290	0.517	0.517	0.517	0.0009	0.904	5
	Kum	1.071	0.770	0.581	0.615	NaN	0.086	0.091	6
	TG	2.640	2.250	0.613	0.629	0.729	0.056	0.126	
0.850	Beta	31.571	6.230	0.835	0.841	0.854	0.004	0.610	
	Kum	16.695	12.051	0.832	0.841	0.859	0.004	0.643	8
	TG	195.800	231.444	0.845	0.844	0.842	0.003	0.594	
	TW	0.881	498.329	0.468	0.455	0.881	0.086	0.090	9
0.900	Beta	17.400	2.488	0.875	0.888	0.917	0.005	0.552	
	Kum	13.055	2.908	0.874	0.888	0.918	0.005	0.566	11
	TG	102.490	112.656	0.883	0.889	0.901	0.005	0.503	
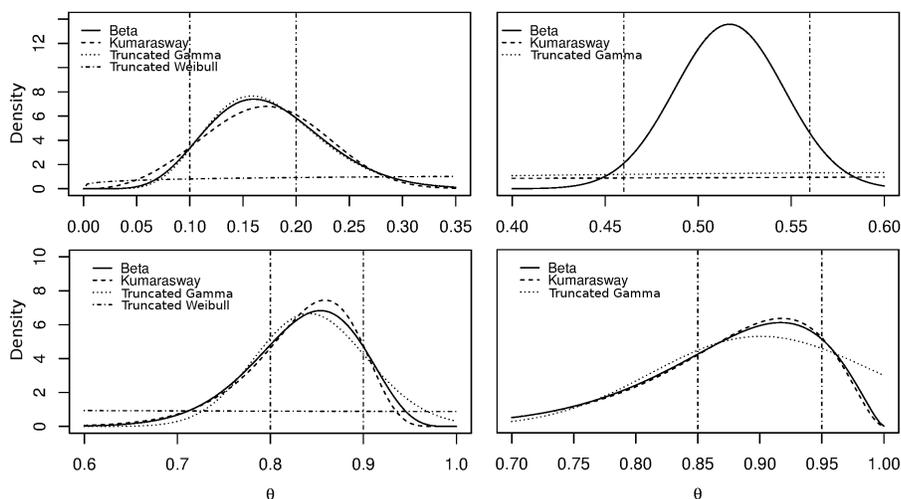
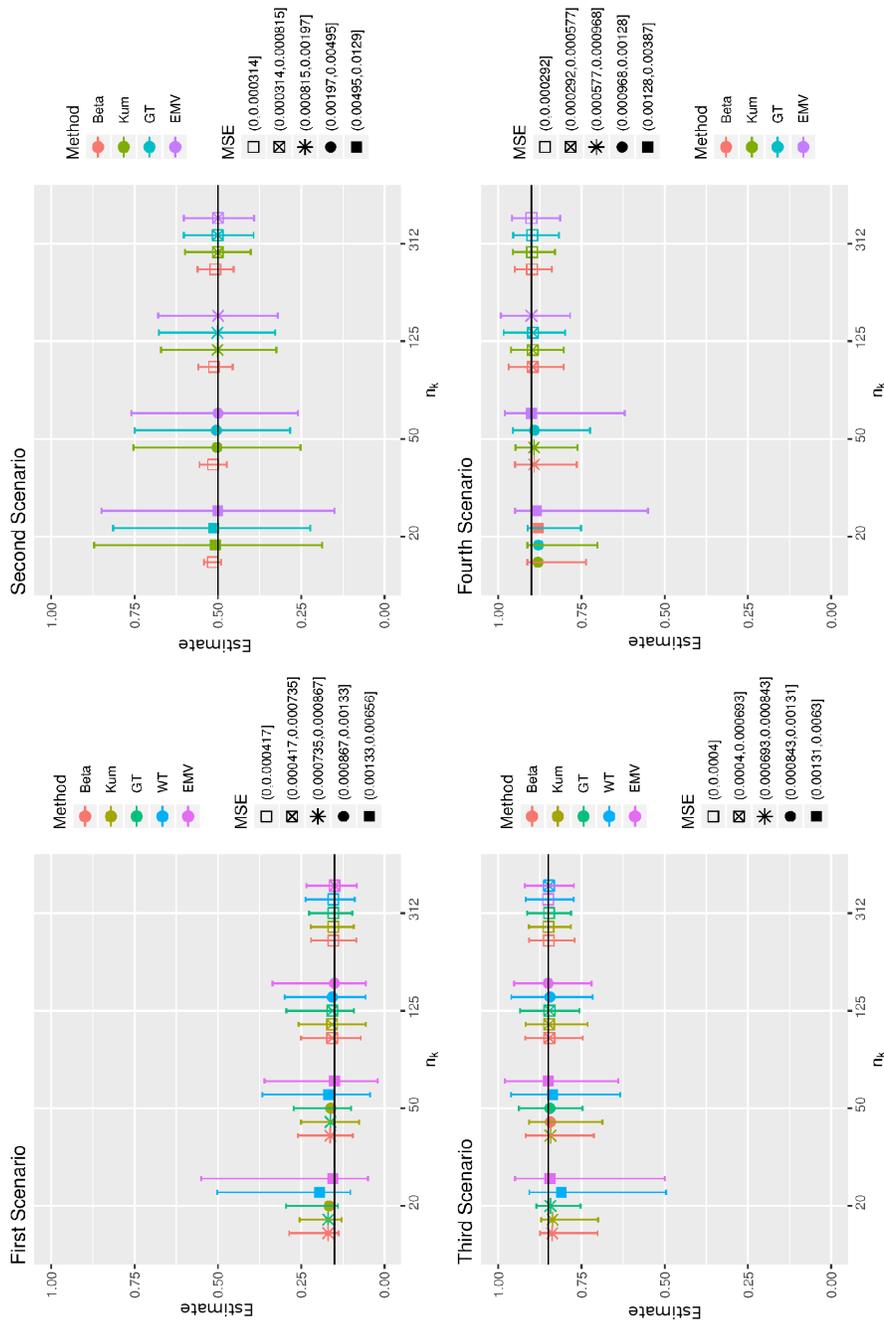FIGURE 3: Densities elicited for the four scenarios of the generating process.

FIGURE 4: Precision and accuracy of estimates using the Maximum Likelihood and Bayesian approach through prior distributions obtained by the proposed method.

The results are shown in Figure 4, which is complemented in Appendix B. The precision-defined as the amplitude of the interval formed by the minimum and maximum values of the estimates-was greater in the first, third and fourth scenarios of the subjective approach when considering the Beta, Kumaraswamy and TG distributions obtained by the proposed elicitation method rather than from the classical approach. In the second scenario, the Beta distribution was the only one that generated considerably greater precision than the classical approach and other prior distributions obtained. This was due to the fact that its variabilities were greater, and the probability contained in the specialist's interval was less than 12.6%. The same was true for the reason why the prior TW had poor precision and accuracy (average of estimates) in the first and third scenarios. As expected, in all scenarios and for both approaches, the increase in sample size generated better precision and accuracy. The MSE was calculated and represented with a symbol on each vertical line of Figure 4 in order to compare the dispersion of the estimates generated by the prior distribution obtained and the classical approach.

The Beta distribution obtained the lowest MSE in all scenarios, followed by Kumarswamy and TG in the first and third scenarios, respectively (Table 5).

5. Aplication: Prevalences in an Indogenous Community

An epidemiological study to estimate the prevalence of three sexually transmitted infections in a Colombian indigenous community was conducted by the University of Antioquia in the eastern region of Colombian. The study enrolled 295 individuals and three HIV infection and eight Syphilis infection cases were observed. The objective of this study was to estimate the prevalence of each disease. This data were used before by Tovar (2012) to obtain the hyperparameter values for a prior Beta distribution using the fixed interval and the Chebyshev inequality. For HIV infection, the author used information presented by the Ministerio de la Protección Social (2010) supported by the criterion of a specialist in the subject, who indicated that the prevalence of this infection should not exceed 1%. According to the Ministerio de la Protección Social, the prevalence of HIV in the population with age between 14 and 49 years was 0.22% and for the entire Colombian population was 0.59%. To illustrate the proposed method, we expressed the prevalences as quotients generating a range of quantiles $(0.0022 = 2.2/1000 \approx 2/1000)$ and $(0.0059 = 5.9/1000 \approx 6/1000)$ for the number of individuals infected with HIV in a hypothetical sample of 1000. With this information, it is possible to define a system of non-linear equations whose roots are the hyperparameter values of a selected prior distribution.

$$\begin{aligned} \widehat{I}_L(x_{(1000),1} = 2 \mid \phi = (a,b)) - L(0.0022 \mid x_{(1000),1}) = 0 \\ \widehat{I}_L(x_{(1000),2} = 6 \mid \phi = (a,b)) - L(0.0059 \mid x_{(1000),2}) = 0 \end{aligned} \tag{17}$$

For the infection of Syphilis, the specialist provided an interval of values for the prevalence $(0.01, 0.03)$ and the system of non-linear equations for this case was:

$$\begin{aligned}
\widehat{I}_L(x_{(100),1} = 1 \mid \phi = (a,b)) - L(0.01 \mid x_{(100),1}) = 0 \\
\widehat{I}_L(x_{(100),2} = 3 \mid \phi = (a,b)) - L(0.03 \mid x_{(100),2}) = 0
\end{aligned} \tag{18}$$

We did not obtain prior distribution for HIV infection, because the prior interval obtained for this infection was in the limits of the parametric space (very close to zero) and the amplitude was too small, which made difficult to find a region D from the plane ab containing at least one point of intersection between the surfaces of the system (17) and the plane ab. For the Syphilis case, we obtained the prior distributions. In Table 6 the characteristics of these prior distributions are showed and the obtained hyperparameter values are compared with those obtained using the method proposed by Tovar (2012). The Beta prior distribution obtained with the proposed method showed the highest density and the highest contained probability within the interval suggested by the specialist. See the upper part of Figure 5. The posterior densities were obtained with the help of the LM. The posterior densities and the binomial likelihood of the observed data set are showed in Figure 5.

TABLE 6: Descriptive measures of elicited prior distributions for syphilis infection. Pr represents $P[\theta \in (\theta_1, \theta_2)]$.

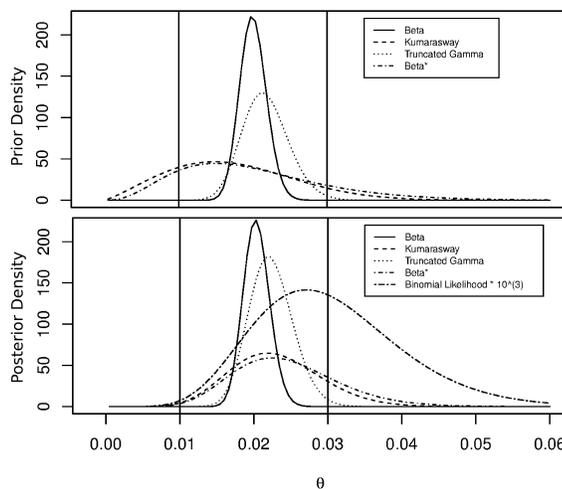| Prior	ϕ^*		Measures of Central Tendency			Var	Pr
	a	b	Mean	Median	Mode		
Beta	122.244	5976.849	0.020	0.020	0.020	3×10^{-6}	0.999
Kum	2.194	5563.686	0.017	0.017	0.015	7×10^{-5}	0.717
TG	49.117	2263.062	0.029	0.022	0.021	9×10^{-5}	0.993
Beta*	3.900	191.100	0.020	0.018	0.015	1×10^{-4}	0.704

FIGURE 5: Densities for prevalence of syphilis infection.

In the infection by Syphilis, we obtained the estimates of the prevalence parameter, using the maximum likelihood and the Bayes estimates, assuming as prior distributions those obtained applying the proposed method and a Beta prior distribution obtained using the Tovar (2012) method (Beta*). Some important indexes as standard deviation, the prior predictive probability and the Bayes Factor to select among prior models and 95% credibility intervals were computed too. See Table 7.

TABLE 7: Posterior estimates obtained using classical and subjective approaches

Method	i		$E[\theta]$	$V[\theta]$	$\widehat{I}_L(x \mid \phi^*)$	BF_{ij}	$CI(95\%)$
Classic		MLE	0.027				
Bayesian	1	Beta	0.020	3×10^{-6}	0.099	BF_{12}=1.518	(0.017, 0.024)
	2	Kum	0.023	4×10^{-5}	0.065	BF_{23}=0.589	(0.012, 0.038)
	3	TG	0.022	1×10^{-5}	0.111	BF_{34}=1.587	(0.017, 0.028)
	4	Beta*	0.024	5×10^{-5}	0.070	BF_{41}=0.704	(0.013, 0.039)
$BF_{13} = 0.895, BF_{24} = 0.935$							

6. Discussion

The proposed method of elicitation in the same way that those developed by Chaloner & Duncan (1983), Tovar (2012), Sindhu et al. (2013) and Flórez & Correa (2015), to assume any prior distribution that has the same domain of the parameter of interest and that is adequate for expressing the specialist's information. In the model this fact makes possible to reflect the natural behavior of the random quantity, θ, complying with the coherence principle; otherwise it will not be possible to obtain the values of the hyperparameters. This result allow us to conclude that Truncated Weibull is not a good choice for a prior distribution when we have a Binomial likelihood. An interesting feature of the method is that the resulting prior distribution elicited is not unique. Therefore the specialist can choose from a wide variety of prior distribution.

The proposed method is similar that developed by Penha (2014). She used the Laplace Method to obtain the hyperparameters of a bivariate Gamma distribution used as prior for a Weibull likelihood. She considers only four observations (one for parameter) but our approach works with a sample of n observations using a nonlinear system equations. This fact improves the error of approximation $O(n^{-1})$. The initial point for the Newton-Raphson method was identified through the contour lines associated with the surfaces of the nonlinear equations system (5). Questions were asked using hypothetical samples to evaluate information from the specialist, who was allowed to select the prior distribution based on the probability contained in the evaluated interval and the asymmetry of the density obtained. Several scenarios for the specialist's beliefs were considered. This made it possible to determine the behavior of the precision and accuracy of the posterior estimates generated by each family of prior distributions.

The elicitation methods described by Flórez & Correa (2015) used only the prior distribution characteristics involved; i.e., its moments, quantils and

probabilities. This allows simulation procedures such as those used by Moala & Penha (2016), which are adequate to evaluate the effectiveness of the foregoing methods but not for the one developed herein. This is due to the fact that the method used by Moala considers the known hyperparameters of the prior distribution that accurately reflects the specialist's knowledge. Then the aforementioned elicitation methods were used, assuming that the evaluated mean and probabilities by the specialist have known errors. As expected, the elicitation methods turn out to be quite precise. Moala reports three specific methods. While the proposed method uses information regarding the likelihood function and the prior distribution considered representative of the specialist's knowledge.

7. Conclusions

Two necessary, but not sufficient conditions were found for the existence of a solution point of the system of nonlinear equations (5). The first is that according to the prior distribution chosen and the hypothetical sample used to evaluate the specialist's information, the probabilities α_j assigned by the expert must not exceed the prior predictive pseudo-probabilities that would be obtained using the prior chosen. If this occurs, the surfaces of the system move away from the *ab* plane and prevent finding a solution point; that is, a vector of hyperparameters that represents the specialist's behavior. The second condition is that the interval proposed by the specialist should not be too width, because the surfaces of the system will be too far apart and could not be intercepted on plane *ab*.

Of the four prior distributions considered, the TW was the one that had the lowest prior predictive pseudo-probability, which prevented obtaining values for the hyperparameters in two of the scenarios and consequently, posterior estimates. In general there was no greater difference between the averages of the posterior estimates generated from the prior distributions obtained for each scenario.

With the proposed methodology is possible to obtain a region D in the plane *ab* where the surfaces of the system of equations is defined using the specialist's information. The obtained level curves are plotted on D and with the Newton-Raphson's method to obtain a vector ϕ^* that represents the additional beliefs about the generating process.

For the case applied to the Colombian indigenous community the prior distribution for the prevalence of HIV infection was not obtained due to the computational cost of finding a D region that contain roots for the system of equations. One way to address this problem is to find a maximum point of each surface and build the region D around these points.

Appendix A. Approximation of the Posterior Expressions

Table A1: Approximation of distribution and r-th posterior moment, considering Bernoulli trials and Beta prior distribution.

Posterior

Distribution $\pi(\theta \mid x_{(n)})$

Expre.
$$\frac{\theta^{x+a-1}(1-\theta)^{n+b-x-1}}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n+b-x-1}d\theta}$$

Appro.
$$\frac{\theta^{c_1^0}(1-\theta)^{c_2}}{(2\pi)^{1/2} \mid c_3^0 \mid^{c_4^0} c_1^{0^{c_1^0}} c_2^{c_2} \mid c_1^0 c_2 \mid^{1/2}}$$
$$\begin{aligned} c_1^0 &= x+a-1 \\ c_2 &= n+b-x-1 \\ c_3^0 &= n+a+b-2 \\ c_4^0 &= 1/2-a-b-n \end{aligned}$$

---

r-th moment centered on zero $E[\theta^r \mid x_{(n)}]$

Expre.
$$\frac{\int_\Omega \theta^{r+x+a-1}(1-\theta)^{n+b-x-1}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n+b-x-1}d\theta}$$

Appro.
$$\frac{\mid c_3^r \mid^{c_4^r} c_1^{r^{c_1^r}} c_2^{c_2} \mid c_1^r c_2 \mid^{1/2}}{\mid c_3^0 \mid^{c_4^0} c_1^{0^{c_1^0}} c_2^{c_2} \mid c_1^0 c_2 \mid^{1/2}}$$
$$\begin{aligned} c_1^r &= r+x+a-1 \\ c_3^r &= r+n+a+b-2 \\ c_4^r &= 1/2-r-a-b-n \end{aligned}$$

Mean $E[\theta \mid x_{(n)}]$

Expre.
$$\frac{\int_\Omega \theta^{x+a}(1-\theta)^{n+b-x-1}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n+b-x-1}d\theta}$$

Appro.
$$\frac{\mid c_3^1 \mid^{c_4^1} c_1^{1^{c_1^1}} c_2^{c_2} \mid c_1^1 c_2 \mid^{1/2}}{\mid c_3^0 \mid^{c_4^0} c_1^{0^{c_1^0}} c_2^{c_2} \mid c_1^0 c_2 \mid^{1/2}}$$
$$\begin{aligned} c_1^1 &= x+a \\ c_3^1 &= n+a+b-1 \\ c_4^1 &= -1/2-a-b-n \end{aligned}$$

Variance $V[\theta \mid x_{(n)}]$

Expre.
$$\frac{\int_\Omega \theta^{1+x+a}(1-\theta)^{n+b-x-1}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n+b-x-1}d\theta} - \left(\frac{\int_\Omega \theta^{x+a}(1-\theta)^{n+b-x-1}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n+b-x-1}d\theta}\right)^2$$

Appro.
$$\frac{\mid c_3^2 \mid^{c_4^2} c_1^{2^{c_1^2}} c_2^{c_2} \mid c_1^2 c_2 \mid^{1/2}}{\mid c_3^0 \mid^{c_4^0} c_1^{0^{c_1^0}} c_2^{c_2} \mid c_1^0 c_2 \mid^{1/2}} - \left(\frac{\mid c_3^1 \mid^{c_4^1} c_1^{1^{c_1^1}} c_2^{c_2} \mid c_1^1 c_2 \mid^{1/2}}{\mid c_3^0 \mid^{c_4^0} c_1^{0^{c_1^0}} c_2^{c_2} \mid c_1^0 c_2 \mid^{1/2}}\right)^2$$
$$\begin{aligned} c_1^2 &= 1+x+a \\ c_3^2 &= n+a+b \\ c_4^2 &= -3/2-a-b-n \end{aligned}$$

TABLE A2: Approximation of distribution and r-th posterior moment, considering Bernoulli trials and Kumaraswamy prior distribution.

Posterior		
Distribution $\pi(\theta \mid x_{(n)})$	Expre.	$\dfrac{\theta^{x+a-1}(1-\theta)^{n-x}(1-\theta^a)^{b-1}}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}(1-\theta^a)^{b-1}d\theta}$
	Appro.	$\dfrac{\theta^{c_1^0}(1-\theta)^{c_2}(1-\theta^a)c_3}{(2\pi)^{1/2}\left|\dfrac{c_1^0}{\widehat{\theta}_D^2}+\dfrac{c_2}{(1-\widehat{\theta}_D)^2}+c_3 a\widehat{\theta}_D^{a-2}\dfrac{(a-1)(1-\widehat{\theta}_D)^a+a\widehat{\theta}_D^a}{(1-\widehat{\theta}_D^a)^2}\right|^{-1/2}\widehat{\theta}_D^{c_1^0}(1-\widehat{\theta}_D)^{c_2}(1-\widehat{\theta}_D^a)^{c_3}}$
		$\widehat{\theta}_D$ is the root of $\dfrac{c_1^0}{\theta}-\dfrac{c_2}{1-\theta}-\dfrac{c_3 a\theta^{a-1}}{1-\theta^a}$ $\quad\begin{aligned}c_1^0 &= x+a-1\\ c_2 &= n-x\\ c_3 &= b-1\end{aligned}$
r-th moment centered on zero $E[\theta^r \mid x_{(n)}]$	Expre.	$\dfrac{\int_\Omega \theta^{r+x+a-1}(1-\theta)^{n-x}(1-\theta^a)^{b-1}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}(1-\theta^a)^{b-1}d\theta}$
	Appro.	$\dfrac{\left|\dfrac{c_1^r}{\widehat{\theta}_N^2}+\dfrac{c_2}{(1-\widehat{\theta}_N)^2}+c_3 a\widehat{\theta}_N^{a-2}\dfrac{(a-1)(1-\widehat{\theta}_N)^a+a\widehat{\theta}_N^a}{(1-\widehat{\theta}_N^a)^2}\right|^{-1/2}\widehat{\theta}_N^{c_1^r}(1-\widehat{\theta}_N)^{c_2}(1-\widehat{\theta}_N^a)^{c_3}}{\left|\dfrac{c_1^0}{\widehat{\theta}_D^2}+\dfrac{c_2}{(1-\widehat{\theta}_D)^2}+c_3 a\widehat{\theta}_D^{a-2}\dfrac{(a-1)(1-\widehat{\theta}_D)^a+a\widehat{\theta}_D^a}{(1-\widehat{\theta}_D^a)^2}\right|^{-1/2}\widehat{\theta}_D^{c_1^0}(1-\widehat{\theta}_D)^{c_2}(1-\widehat{\theta}_D^a)^{c_3}}$
		$\widehat{\theta}_N$ is the root of $-nh_r'(\theta)=\dfrac{c_1^r}{\theta}-\dfrac{c_2}{1-\theta}-\dfrac{c_3 a\theta^{a-1}}{1-\theta^a}$ $\qquad c_1^r = r+x+a-1$
Mean $E[\theta \mid x_{(n)}]$	Appro.	Using the approximated of the r-th posterior moment with $r=1$, the approximation of the posterior mean is obtained with $c_1^1 = x+a$ and $\widehat{\theta}_N$ as the root of $-nh_1'(\theta)$.
Variance $V[\theta \mid x_{(n)}]$	Appro.	It is enough know the approximation of the second posterior moment.

TABLE A3: Approximation of distribution and r-th posterior moment, considering Bernoulli trials and Truncated Gamma prior distribution.

Posterior		
Distribution $\pi(\theta \mid x_{(n)})$	Expre.	$\dfrac{\theta^{x+a-1}(1-\theta)^{n-x}e^{-b\theta}}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}e^{-b\theta}d\theta}$
	Appro.	$\dfrac{\theta^{c_1^0}(1-\theta)^{c_2}e^{-b\theta}}{(2\pi)^{1/2}\left\|\frac{c_1^0}{\widehat{\theta}_D^2}+\frac{c_2}{(1-\widehat{\theta}_D)^2}\right\|^{-1/2}\widehat{\theta}_D^{c_1^0}(1-\widehat{\theta}_D)^{c_2}e^{-b\widehat{\theta}_D}}$ $\begin{aligned} c_1^0 &= x+a-1 \\ c_2 &= n-x \\ c_3^0 &= n+a+b-1 \end{aligned}$ $\widehat{\theta}_D = \dfrac{c_3^0 \pm \sqrt{(c_3^0)^2 - 4bc_1^0}}{2b}$
r-th moment centered on zero $E[\theta^r \mid x_{(n)}]$	Expre.	$\dfrac{\int_\Omega \theta^{r+x+a-1}(1-\theta)^{n-x}e^{-b\theta}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}e^{-b\theta}d\theta}$
	Appro.	$\dfrac{\left\|\frac{c_1^r}{\widehat{\theta}_N^2}+\frac{c_2}{(1-\widehat{\theta}_N)^2}\right\|^{-1/2}\widehat{\theta}_N^{c_1^r}(1-\widehat{\theta}_N)^{c_2}}{\left\|\frac{c_1^0}{\widehat{\theta}_D^2}+\frac{c_2}{(1-\widehat{\theta}_D)^2}\right\|^{-1/2}\widehat{\theta}_D^{c_1^0}(1-\widehat{\theta}_D)^{c_2}}e^{-b(\widehat{\theta}_N+\widehat{\theta}_D)}$ $\begin{aligned} c_1^r &= r+x+a-1 \\ c_3^r &= r+n+a+b-1 \end{aligned}$ $\widehat{\theta}_N^r = \dfrac{c_3^r \pm \sqrt{(c_3^r)^2 - 4bc_1^r}}{2b}$
Mean $E[\theta \mid x_{(n)}]$	Appro.	Using the approximated of the r-th posterior moment with $r=1$, the approximation of the posterior mean is obtained with $c_1^1 = x+a$, $c_3^1 = n+a+b$ y $\widehat{\theta}_N^1$ as the root of $-nh_1'(\theta)$.
Variance $V[\theta \mid x_{(n)}]$	Appro.	It is enough know the approximation of the second posterior moment.

TABLE A4: Approximation of distribution and r-th posterior moment, considering Bernoulli trials and Truncated Weibull prior distribution.

Posterior		
Distribution $\pi(\theta \mid x_{(n)})$	Expre.	$\dfrac{\theta^{x+a-1}(1-\theta)^{n-x}e^{-(\theta/b)^a}}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}e^{-(\theta/b)^a}d\theta}$
	Appro.	$\dfrac{\theta^{c_1^0}(1-\theta)^{c_2}e^{-(\theta/b)^a}}{(2\pi)^{1/2}\left\lvert \dfrac{c_1^0}{\widehat\theta_D^2} + \dfrac{c_2}{(1-\widehat\theta_D)^2} + \dfrac{a(a-1)\widehat\theta^{a-2}}{b^a}\right\rvert^{-1/2}\widehat\theta_D^{c_1^0}(1-\widehat\theta_D)^{c_2}e^{-(\widehat\theta_D/b)^a}}$ $\widehat\theta_D$ is the root of $\dfrac{c_1^0}{\theta} - \dfrac{c_2}{1-\theta} - \left(\dfrac{a}{b}\right)\left(\dfrac{\theta}{b}\right)^{a-1}$; $\quad \begin{aligned} c_1^0 &= x+a-1 \\ c_2 &= n-x \end{aligned}$
r-th moment centered on zero $E[\theta^r \mid x_{(n)}]$	Expre.	$\dfrac{\int_\Omega \theta^{r+x+a-1}(1-\theta)^{n-x}e^{-(\theta/b)^a}d\theta}{\int_\Omega \theta^{x+a-1}(1-\theta)^{n-x}e^{-(\theta/b)^a}d\theta}$
	Appro.	$\dfrac{\left\lvert \dfrac{c_1^r}{\widehat\theta_N^2} + \dfrac{c_2}{(1-\widehat\theta_N)^2} + \dfrac{a(a-1)\widehat\theta^{a-2}}{b^a}\right\rvert^{-1/2}\widehat\theta_N^{c_1^r}(1-\widehat\theta_N)^{c_2}}{\left\lvert \dfrac{c_1^0}{\widehat\theta_D^2} + \dfrac{c_2}{(1-\widehat\theta_D)^2} + \dfrac{a(a-1)\widehat\theta^{a-2}}{b^a}\right\rvert^{-1/2}\widehat\theta_D^{c_1^0}(1-\widehat\theta_D)^{c_2}} e^{-b^{-a}(\widehat\theta_N^a+\widehat\theta_D^a)}$ $\widehat\theta_N$ is the root of $-nh_r'(\theta) = \dfrac{c_1^r}{\theta} - \dfrac{c_2}{1-\theta} - \left(\dfrac{a}{b}\right)\left(\dfrac{\theta}{b}\right)^{a-1}$; $\quad c_1^r = r+x+a-1$
Mean $E[\theta \mid x_{(n)}]$	Appro.	From the approximation of the r-th posterior moment, with $r = 1$, the approximation of the posterior mean is obtained with $c_1^1 = x+a$ y $\widehat\theta_N$ the root of $-nh_1'(\theta)$.
Variance $V[\theta \mid x_{(n)}]$	Appro.	It is enough know the approximation of the second posterior moment.

Appendix B. Results Simulation

The notation used is the following: $\widehat{\theta}^k_{pt,s}$ and $\widehat{\sigma^2}^k_{pt,s}$ are respectively the s-th estimation of the mean and variance posterior; the prior predictive probability for the s-th sample of size n_k is denoted as $\widehat{I}^k_s$; the s-th maximum likelihood estimation for sample n_k is denoted as $\widehat{\theta}^k_{M,s}$.

TABLE B1: Scenario of the generating process $\theta = 0.1500$ for four sample sizes.

θ	Prior	n_k	$\overline{\widehat{\theta}}^k_{pt}(\min \widehat{\theta}^k_{pt,s}, \max \widehat{\theta}^k_{pt,s})$	$\overline{\overline{\widehat{\sigma^2}}}^k_{pt}(\min \widehat{\sigma^2}^k_{pt,s}, \max \widehat{\sigma^2}^k_{pt,s})$	$\overline{\widehat{I}}^k(\min \widehat{I}^k_s, \max \widehat{I}^k_s)$	$\overline{\widehat{\theta}}^k_M(\min \widehat{\theta}^k_{M,s}, \max \widehat{\theta}^k_{M,s})$
		20	0.1690 (0.1377, 0.2861)	0.0020 (0.0017, 0.0030)	0.1627 (0.0012, 0.2016)	0.1554 (0.0500, 0.5500)
	Beta	50	0.1624 (0.0953, 0.2595)	0.0014 (0.0009, 0.0020)	0.0883 (0.0077, 0.1060)	0.1509 (0.0200, 0.3400)
		125	0.1566 (0.0712, 0.2510)	0.0008 (0.0004, 0.0011)	0.0447 (0.0015, 0.0505)	0.1498 (0.0320, 0.2800)
		312	0.1532 (0.0842, 0.2206)	0.0004 (0.0002, 0.0005)	0.0205 (0.0041, 0.0221)	0.1499 (0.0705, 0.2276)
		20	0.1663 (0.1285, 0.2546)	0.0022 (0.0018, 0.0023)	0.1666 (0.0030, 0.2024)	0.1551 (0.0500, 0.5000)
	Kum	50	0.1610 (0.0760, 0.2507)	0.0015 (0.0011, 0.0015)	0.0870 (0.0116, 0.1032)	0.1495 (0.0200, 0.3400)
		125	0.1571 (0.0564, 0.2572)	0.0008 (0.0004, 0.0009)	0.0424 (0.0043, 0.0477)	0.1498 (0.0320, 0.2960)
0.1500		312	0.1535 (0.0919, 0.2211)	0.0004 (0.0003, 0.0005)	0.0188 (0.0082, 0.0206)	0.1497 (0.0833, 0.2276)
		20	0.1688 (0.1394, 0.2952)	0.0020 (0.0016, 0.0038)	0.1625 (0.0013, 0.2010)	0.1543 (0.0500, 0.5500)
	TG	50	0.1620 (0.0997, 0.2728)	0.0013 (0.0008, 0.0022)	0.0880 (0.0068, 0.1070)	0.1499 (0.0200, 0.3600)
		125	0.1565 (0.0915, 0.2943)	0.0007 (0.0004, 0.0013)	0.0454 (0.0033, 0.0515)	0.1497 (0.0560, 0.3360)
		312	0.1530 (0.0966, 0.2259)	0.0004 (0.0002, 0.0005)	0.0211 (0.0065, 0.0228)	0.1498 (0.0833, 0.2340)
		20	0.1951 (0.1021, 0.5015)	0.0065 (0.0038, 0.0107)	0.0428 (0.0355, 0.0522)	0.1561 (0.0500, 0.5000)
	TW	50	0.1673 (0.0437, 0.3670)	0.0026 (0.0008, 0.0044)	0.0173 (0.0118, 0.0203)	0.1507 (0.0200, 0.3600)
		125	0.1568 (0.0568, 0.3001)	0.1568 (0.0568, 0.3001)	0.0010 (0.0004, 0.0016)	0.0070 (0.0056, 0.0079)
		312	0.1529 (0.0898, 0.2361)	0.0004 (0.0003, 0.0006)	0.0028 (0.0025, 0.0031)	0.1502 (0.0865, 0.2340)

TABLE B2: Scenario of the generating process, $\theta = 0.5000$ and $\theta = 0.8500$, for four sample sizes.

θ	Prior	n_k	$\overline{\hat{\theta}}_{pt}^k (\min \hat{\theta}_{pt,s}^k, \max \hat{\theta}_{pt,s}^k)$	$\overline{\widehat{\sigma^2}}_{pt}^k (\min \widehat{\sigma^2}_{pt,s}^k, \max \widehat{\sigma^2}_{pt,s}^k)$	$\overline{\hat{I}}^k (\min \hat{I}_s^k, \max \hat{I}_s^k)$	$\overline{\hat{\theta}}_M^k (\min \hat{\theta}_{M,s}^k, \max \hat{\theta}_{M,s}^k)$
0.5000	Beta	20	0.5157 (0.4899, 0.5414)	0.0008 (0.0008, 0.0008)	0.1231 (0.0002, 0.1690)	0.5015 (0.1000, 0.9000)
		50	0.5142 (0.4731, 0.5554)	0.0007 (0.0007, 0.0007)	0.0754 (0.0000, 0.1039)	0.4999 (0.2200, 0.7800)
		125	0.5116 (0.4551, 0.5586)	0.0006 (0.0006, 0.0006)	0.0444 (0.0000, 0.0596)	0.4997 (0.3120, 0.6560)
		312	0.5081 (0.4533, 0.5612)	0.0004 (0.0004, 0.0004)	0.0243 (0.0003, 0.0314)	0.5002 (0.3942, 0.6026)
	Kum	20	0.5078 (0.1870, 0.8712)	0.0105 (0.0050, 0.0109)	0.0455 (0.0368, 0.0635)	0.5021 (0.1500, 0.9000)
		50	0.5029 (0.2521, 0.7532)	0.0046 (0.0036, 0.0047)	0.0182 (0.0157, 0.0220)	0.5002 (0.2400, 0.7600)
		125	0.5012 (0.3238, 0.6706)	0.0019 (0.0017, 0.0020)	0.0073 (0.0066, 0.0082)	0.5001 (0.3200, 0.6720)
		312	0.5004 (0.4017, 0.5992)	0.0008 (0.0008, 0.0008)	0.0029 (0.0028, 0.0031)	0.4999 (0.4006, 0.5994)
	TG	20	0.5126 (0.2232, 0.8153)	0.0098 (0.0065, 0.0101)	0.0580 (0.0233, 0.0670)	0.5007 (0.1500, 0.8500)
		50	0.5048 (0.2832, 0.7497)	0.0045 (0.0035, 0.0046)	0.0242 (0.0148, 0.0273)	0.4997 (0.2600, 0.7600)
		125	0.5015 (0.3276, 0.6775)	0.0019 (0.0017, 0.0019)	0.0098 (0.0072, 0.0109)	0.4994 (0.3200, 0.6800)
		312	0.5005 (0.3932, 0.6023)	0.0008 (0.0008, 0.0008)	0.0040 (0.0034, 0.0043)	0.4997 (0.3910, 0.6026)
0.8500	Beta	20	0.8381 (0.7017, 0.8747)	0.0023 (0.0019, 0.0036)	0.1633 (0.0013, 0.1973)	0.8443 (0.4500, 0.9500)
		50	0.8434 (0.7126, 0.9175)	0.0015 (0.0009, 0.0023)	0.0872 (0.0049, 0.1035)	0.8497 (0.6200, 0.9800)
		125	0.8467 (0.7467, 0.9187)	0.0008 (0.0005, 0.0012)	0.0434 (0.0109, 0.0480)	0.8502 (0.7200, 0.9440)
		312	0.8483 (0.7706, 0.9079)	0.0004 (0.0002, 0.0005)	0.0197 (0.0089, 0.0207)	0.8499 (0.7628, 0.9167)
	Kum	20	0.8370 (0.6995, 0.8708)	0.0024 (0.0016, 0.0048)	0.1596 (0.0041, 0.1953)	0.8430 (0.5000, 0.9500)
		50	0.8431 (0.6869, 0.9067)	0.0014 (0.0005, 0.0030)	0.0875 (0.0038, 0.1059)	0.8493 (0.6000, 0.9800)
		125	0.8476 (0.7314, 0.9172)	0.0008 (0.0004, 0.0013)	0.0450 (0.0067, 0.0506)	0.8508 (0.7040, 0.9520)
		312	0.8488 (0.7825, 0.9081)	0.0004 (0.0002, 0.0005)	0.0208 (0.0089, 0.0222)	0.8499 (0.7756, 0.9199)
	TG	20	0.8432 (0.7524, 0.8854)	0.0022 (0.0018, 0.0024)	0.1655 (0.0021, 0.1991)	0.8448 (0.5000, 0.9500)
		50	0.8446 (0.7472, 0.9379)	0.0014 (0.0012, 0.0015)	0.0869 (0.0048, 0.1025)	0.8498 (0.6400, 0.9800)
		125	0.8459 (0.7562, 0.9345)	0.0008 (0.0005, 0.0009)	0.0427 (0.0089, 0.0470)	0.8501 (0.7200, 0.9520)
		312	0.8475 (0.7808, 0.9128)	0.0004 (0.0002, 0.0004)	0.0192 (0.0103, 0.0202)	0.8499 (0.7724, 0.9199)
	TW	20	0.8104 (0.4964, 0.9065)	0.0065 (0.0038, 0.0109)	0.0436 (0.0410, 0.0474)	0.8445 (0.5000, 0.9500)
		50	0.8367 (0.6335, 0.9611)	0.0025 (0.0007, 0.0044)	0.0177 (0.0163, 0.0185)	0.8509 (0.6400, 0.9800)
		125	0.8446 (0.7162, 0.9605)	0.0010 (0.0003, 0.0016)	0.0072 (0.0069, 0.0073)	0.8503 (0.7200, 0.9680)
		312	0.8478 (0.7738, 0.9172)	0.0004 (0.0002, 0.0006)	0.0029 (0.0028, 0.0029)	0.8501 (0.7756, 0.9199)

TABLE B3: Scenario of the generating process $\theta = 0.9000$ for four sample sizes.

θ	Prior	n_k	$\overline{\widehat{\theta}}{}^k_{pt}(\min \widehat{\theta}^k_{pt,s}, \max \widehat{\theta}^k_{pt,s})$	$\overline{\widehat{\sigma^2}}{}^k_{pt}(\min \widehat{\sigma^2}{}^k_{pt,s}, \max \widehat{\sigma^2}{}^k_{pt,s})$	$\overline{\widehat{I}}{}^k(\min \widehat{I}^k_s, \max \widehat{I}^k_s)$	$\overline{\widehat{\theta}}{}^k_M(\min \widehat{\theta}^k_{M,s}, \max \widehat{\theta}^k_{M,s})$
		20	0.8798 (0.7366, 0.9120)	0.0026 (0.0020, 0.0047)	0.1795 (0.0125, 0.2134)	0.8858 (0.6000, 0.9500)
	Beta	50	0.8919 (0.7639, 0.9499)	0.0013 (0.0007, 0.0025)	0.0901 (0.0180, 0.1030)	0.8990 (0.7200, 0.9800)
		125	0.8964 (0.8033, 0.9690)	0.0006 (0.0002, 0.0011)	0.0417 (0.0175, 0.0453)	0.8999 (0.7920, 0.9840)
		312	0.8985 (0.8388, 0.9503)	0.0003 (0.0001, 0.0004)	0.0180 (0.0117, 0.0190)	0.9000 (0.8365, 0.9551)
0.9000		20	0.8808 (0.7021, 0.9121)	0.0025 (0.0019, 0.0058)	0.1790 (0.0066, 0.2125)	0.8864 (0.5500, 0.9500)
	Kum	50	0.8925 (0.7623, 0.9483)	0.0013 (0.0007, 0.0027)	0.0910 (0.0178, 0.1047)	0.8992 (0.7200, 0.9800)
		125	0.8967 (0.8035, 0.9609)	0.0006 (0.0002, 0.0011)	0.0426 (0.0169, 0.0467)	0.8999 (0.7920, 0.9760)
		312	0.8985 (0.8300, 0.9559)	0.0003 (0.0001, 0.0004)	0.0184 (0.0103, 0.0196)	0.8998 (0.8269, 0.9615)
		20	0.8794 (0.7509, 0.9118)	0.0024 (0.0021, 0.0027)	0.1839 (0.0039, 0.2268)	0.8856 (0.5500, 0.9500)
	TG	50	0.8916 (0.7247, 0.9562)	0.0013 (0.0005, 0.0019)	0.0876 (0.0020, 0.0964)	0.9003 (0.6200, 0.9800)
		125	0.8955 (0.7986, 0.9838)	0.0007 (0.0002, 0.0010)	0.0388 (0.0166, 0.0405)	0.9003 (0.7840, 0.9920)
		312	0.8978 (0.8177, 0.9546)	0.0003 (0.0001, 0.0004)	0.0163 (0.0100, 0.0166)	0.9000 (0.8141, 0.9583)

References

Azevedo-Filho, A. & Shachter, R. D. (1994), Laplace's method approximations for probabilistic inference in belief networks with continuous variables, *in* R. Lopez & D. Poole, eds, 'Uncertainty in Artificial Intelligence : Proceedings of the Tenth Conference', Morgan Kaufmann Publishers Inc., San Francisco, CA, pp. 28–36.

Bruijn, N. G. D. (1961), *Asymptotic Methods in Analysis*, 2 edn, Courier Corporation, Ámsterdam.

Chaloner, K. M. & Duncan, G. T. (1983), 'Assessment of a beta prior distribution: PM elicitation', *Journal of the Royal Statistical Society: Series D (The Statistician)* **32**(1-2), 174–180.

Dennis-Jr, J. E. & Schnabel, R. B. (1996), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Siam.

Erdélyi, A. (1956), *Asymptotic expansions*, number 3, Courier Corporation.

Flórez, A. & Correa, J. (2015), Elicitación de una distribución subjetiva del vector de parámetros π de la distribución multinomial, Tesis de maestría, Universidad Nacional de Colombia, Colombia.

Fowler & Floyd, J. (1995), *Improving Survey Questions: Design and Evaluation*, Vol. 38, Sage.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. (2005), 'Statistical methods for eliciting probability distributions', *Journal of the American Statistical Association* **100**(470), 680–701.

Hogarth, R. M. (1975), 'Cognitive processes and the assessment of subjective probability distributions', *Journal of the American Statistical Association* **70**(350), 271–289.

Hogarth, R. M. (1987), *Judgement and choice: The psychology of decision*, 2 edn.

Kadane, J. B. & Winkler, R. L. (1988), 'Separating probability elicitation from utilities', *Journal of the American Statistical Association* **83**(402), 357–363.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the american statistical association* **90**(430), 773–795.

Laplace, P. S. (1773), 'Memoir on the probability of the causes of events', *Statistical Science* **1**(3), 364–378.

Moala, F. A. & Penha, D. L. (2016), 'Elicitation methods for beta prior distribution', *Revista Brasileira de Biometria* **34**(1), 49–62.

Murphy, A. H. & Winkler, R. L. (1974), 'Credible interval temperature forecasting: some experimental results', *Monthly Weather Review* **102**(11), 784–794.

Penha, D. L. (2014), Inferência bayesiana não-paramétrica para elicitação da função de contabilidade. Universidade Estadual Paulista (UNESP).

Sindhu, T. N., Feroze, N. & Aslam, M. (2013), 'Bayesian analysis of the kumaraswamy distribution under failure censoring sampling scheme', *International Journal of Advanced Science and Technology* **51**, 39–58.

Tierney, L. & Kadane, J. B. (1986), 'Accurate approximations for posterior moments and marginal densities', *Journal of the American Statistical Association* **81**(393), 82–86.

Tovar, J. R. (2012), 'Eliciting beta prior distributions for binomial sampling', *Revista Brasileira de Biometria* **30**(1), 159–172.

Tversky, A. & Kahneman, D. (1973), 'Availability: A heuristic for judging frequency and probability', *Cognitive Psychology* **2**(5), 207–232.

Tversky, A. & Kahneman, D. (1974), 'Judgment under uncertainty: Heuristics and biases', *Science* **185**(4157), 1124–1131.

Vidal, I. (2014), 'A bayesian analysis of the gumbel distribution: an application to extreme rainfall data', *Stochastic Environmental Research and Risk Assessment* **28**(3), 571–582.

Winkler, R. L. (1967), 'The assessment of prior distributions in bayesian analysis', *Journal of the American Statistical Association* **62**(319), 776–800.