HHS Public Access
Author manuscript
SoftwareX. Author manuscript; available in PMC 2023 June 27.

Published in final edited form as:
SoftwareX. 2023 May ; 22: . doi:10.1016/j.softx.2023.101358.

BayesESS: A tool for quantifying the impact of parametric priors in Bayesian analysis

Jaejoon Song[a,*], **Satoshi Morita**[b], **Ying-Wei Kuo**[c], **J. Jack Lee**[c]

[a]Office of Biostatistics, US Food and Drug Administration, Silver Spring, MD 20993, USA

[b]Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan

[c]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Abstract

Bayesian inference has become an attractive choice for scientists seeking to incorporate prior knowledge into their modeling framework. While the `R` community has been an important contributor in facilitating Bayesian statistical analyses, software to evaluate the impact of prior knowledge to such modeling framework has been lacking. In this article, we present `BayesESS`, a comprehensive, free, and open source `R` package for quantifying the impact of parametric priors in Bayesian analysis. We also introduce an accompanying web-based application for estimating and visualizing Bayesian effective sample size for purposes of conducting or planning Bayesian analyses.

Keywords

Bayesian analysis; Effective sample size; Posterior distribution; R package

Code metadata	
Current code version	Version 0.1.19
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-21-00150
Current executable software	https://cran.r-project.org/web/packages/BayesESS/index.html

*Corresponding author. jaejoon.song@fda.hhs.gov (Jaejoon Song).

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer
This publication reflects the views of the authors and should not be construed to represent FDA's views or policies.

Appendix A. Supplementary data
Supplementary material related to this article can be found online at https://doi.org/10.1016/j.softx.2023.101358.

Code metadata

Legal Code License	GNU General Public License (GPL)
Code versioning system used	git
Software code languages, tools, and services used	R, C++
Compilation requirements, operating environments	Windows, Linux, and macOS; R package dependencies: MCMCpack, LaplacesDemon, Rcpp, dfcrm, MatrixModels, MASS, RcppArmadillo, RcppEigen
If available Link to developer documentation/manual	https://cran.r-project.org/web/packages/BayesESS/BayesESS.pdf
Support email for questions	jaejoonsong@gmail.com

Software metadata

Current software version	Version 0.1.19
Permanent link to executables of this version	R package: https://cran.r-project.org/web/packages/BayesESS/index.html;web-based application https://ibl.mdanderson.org/shinyapps/BayesESS/
Legal Software License	GNU General Public License (GPL)
Computing platforms/Operating Systems	Windows, Linux, and macOS
Installation requirements	R packages: MCMCpack, LaplacesDemon, Rcpp, dfcrm, MatrixModels, MASS, RcppArmadillo, RcppEigen
If available, link to user manual - if formally published include a reference to the publication in the reference list	https://cran.r-project.org/web/packages/BayesESS/BayesESS.pdf
Support email for questions	jaejoonsong@gmail.com

1. Motivation and significance

In Bayesian inference, the current state of knowledge or uncertainty about a parameter is expressed as a probability distribution or, in short, *prior* [1]. Such a state of knowledge is updated by data, resulting in a new distribution or, in short, *posterior*. Inference about the parameter of interest is conducted using the posterior distribution, which is a reflection of the information from both the prior distribution and the data [1].

The open source software community has actively facilitated computational tools for Bayesian analysis, with at least 113 software packages related to Bayesian inference currently reposited in the Comprehensive R Archive Network (CRAN). However, there is a void in software to evaluate the impact of prior knowledge in such a modeling framework.

In the design and analysis of clinical trials under the Bayesian paradigm, it is often of interest to assess the amount of information on the posterior, which is influenced by the selection of the informative prior [2–4]. The impact on the posterior distribution from choosing a certain prior distribution is discussed as the *effective sample size* (ESS) in the statistical literature [4]. The ESS quantifies the level of influence on the posterior distribution posed by the choice of prior distribution in the unit of sample size. The approach has been widely accepted in the biostatistical literature because the quantification into the unit of *sample size* is in accordance with the process of sample size determination when designing a study.

A distinction should be made between the ESS discussed in this article and *effective sample size* discussed in the context of Markov Chain Monte Carlo (MCMC) sampling. While the former (ESS, as defined in this article) is used to quantify and communicate the impact of priors, the latter is used to measure the effectiveness of MCMC samples.

2. Software description

We present BayesESS, a free, open-source, comprehensive R package and a web-based application for quantifying the impact of parametric priors in Bayesian analysis. Our R package can be used for determining the ESS for trivial cases where the closed-form solution exists (e.g., conjugate models such as beta-binomial, gamma-exponential, gamma-Poisson, dirichlet-multinomial, normal–normal model), and also for non-trivial cases when the ESS is determined numerically (e.g., linear regression, logistic regression, or time-to-event model). The R package is available from CRAN at https://cran.r-project.org/web/packages/BayesESS/. An accompanying interactive web-based application is available from https://biostatistics.mdanderson.org/shinyapps/BayesESS/ to allow the user to further explore the impact of parametric priors through visualization. The web-based application can also be used in place of the R package by scientists who are unfamiliar with the R programming language. In this article, we demonstrate the software for estimating the Bayesian ESS for purpose of planning or conducting Bayesian analyses.

2.1. Model description

Whereas parameters are assumed to be fixed quantities in the frequentist framework, they are treated as random variables in the Bayesian perspective [1]. In this section, we introduce some notations to illustrate approaches commonly used in Bayesian inference.

Let $\mathbf{y} = y_1, \ldots, y_n$ be a random sample of size n from a random variable $\mathbf{Y}$. Suppose that we are interested in a parameter $\boldsymbol{\theta}$. Then we denote $p(\mathbf{y} \mid \boldsymbol{\theta})$ as a family of density functions over $\mathbf{y}$, parameterized by the random variable $\boldsymbol{\theta}$. We call this family $p(\mathbf{y} \mid \boldsymbol{\theta})$ a *likelihood* function or likelihood model for the data $\mathbf{y}$, given the model specified by any value of θ. The *prior* distribution that describes the uncertainty about the parameter θ is denoted as $p(\boldsymbol{\theta})$. From these quantities, the objective of Bayesian inference is to obtain the updated uncertainty about the parameter of interest ($\boldsymbol{\theta}$) after observing the data. Such updating is expressed in a density called the *posterior* denoted as $p(\boldsymbol{\theta} \mid \mathbf{y})$. Using the Bayes' rule, the *posterior* can be identified as:

$$p\left(\boldsymbol{\theta} \mid \mathbf{y}\right) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y} \mid \boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'}.$$

2.1.1. Types of priors—

The prior distribution plays a central role in Bayesian analysis. Priors that contain minimal information are referred to as *non-informative* priors (also known as the *reference* or the *objective* priors). *Informative* priors are utilized when it is essential to incorporate existing (e.g., clinical) knowledge in the model (see Appendix 1 for an example of a Bayesian model using an informative prior). Elicitation of informative priors can be based on pure judgment, a mixture of data and judgment, or data alone [1].

2.1.2. An analytical method for determining the ESS—Some simple Bayesian models (see conjugate models described in Appendix 2) offer a closed-form quantification for the impact of a parametric prior [1]. However, many commonly used statistical models such as linear or logistic regression or time-to-event models do not extend a closed-form expression for the ESS [2–4] (see model details described in Appendices 1, 6, and 7).

A technique proposed by Morita et al. [4,5] can be used to determine the ESS under general conditions, even when the posterior distribution is intractable. Assume that we have a random sample of data Y, generated from an underlying distribution with parameter $\boldsymbol{\theta}$. The main idea behind the method of Morita et al. [4] is to compute the distance between the prior distribution of interest (denoted $p(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}})$, where $\widetilde{\boldsymbol{\theta}}$ indicates hyperparameters) and a posterior distribution $\left(q_m(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}_0, \mathbf{y}_m)\right)$ generated using a prior having a negligible amount of information (denoted $q_0(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}_0)$, where $\widetilde{\boldsymbol{\theta}}_0$ indicates hyperparameters). The reference prior $\left(q_0(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}_0)\right)$ is specified with the same mean (and the same correlation structure among parameters if the prior distribution is specified with multiple parameters), while inflating the variance. The posterior $\left(q_m(\boldsymbol{\theta} \mid \widetilde{\theta}_0, \mathbf{y}_m)\right)$ is assumed to have an independently and identically distributed (i.i.d) sample of size m. The ESS is identified as a sample size (m) that minimizes the distance between the prior distribution of interest $\left(p(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}})\right)$ and the posterior $\left(q_m(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}_0, \mathbf{y}_m)\right)$. The ESS provides an intuitive assessment of the impact of prior specification, with reference to the sample size in the observed data. For example, if a prior of interest suggests an ESS of 50 when the observed data to be analyzed has 25 observations, then clearly, this implies that the prior will dominate the posterior inference. In the following section, we illustrate how ESS is identified in a simple Bayesian conjugate model.

2.1.3. Illustration of ESS in a conjugate model (Beta-binomial)—Given a data set with distribution $p(\mathbf{y} \mid \boldsymbol{\theta})$, a family of distributions is said to be *conjugate* to the given distribution if the posterior is in the conjugate family whenever the prior is in the conjugate family, regardless of the observed value of the data [1] (see Appendix 2 for examples of conjugate priors). Such prior elicitation is often used because the posterior is easily tractable this way (i.e., there is no need to compute the integral $\int p(\mathbf{y} \mid \boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'$ in the denominator of Bayes' rule). In this section, we illustrate the definition of ESS in a beta-binomial model.

Suppose that we have data generated from m exchangeable Bernoulli trials $y_1, \ldots, y_m$, where $y_i = 1$ is labeled as a "success" and $y_i = 0$ is labeled as a "failure". Then the number of successes in m trials is represented as a binomial distribution $\sum_i^n Y_i \sim Bin(m, \theta)$, such that:

$$f_m(\mathbf{y}_m \mid \boldsymbol{\theta}) = \binom{m}{y} \theta^y \left(1 - \theta\right)^{m - y}.$$

Suppose that we are interested in using a beta prior (with hyperpriors α and β) for θ. To identify an ESS, a posterior distribution $\left(q_m\left(\theta \mid \widetilde{\theta}_0, \mathbf{y}_m\right)\right)$ can be formulated using a beta prior distribution specified with the same mean and an inflated variance as the following, where c is assumed to be a very large constant.

$$q_0\left(\theta \mid \widetilde{\theta}_0\right) = p\left(\theta \mid \frac{\alpha}{c}, \frac{\beta}{c}\right) = B\left(\frac{\alpha}{c}, \frac{\beta}{c}\right)\theta^{\frac{\alpha}{c} - 1}\left(1 - \theta\right)^{\frac{\beta}{c} - 1} \propto \theta^{\frac{\alpha}{c} - 1}\left(1 - \theta\right)^{\frac{\beta}{c} - 1},$$

Note that the objective of ESS computation is to provide an intuitive assessment of the impact of prior specification, with reference to the sample size (m) in the observed data. Therefore, the ESS is identified as a sample size (m) that minimizes the distance between the prior distribution of interest $\left(p\left(\theta \mid \widetilde{\theta}\right)\right)$ and the posterior $\left(q_m\left(\theta \mid \widetilde{\theta}_0, \mathbf{y}_m\right)\right)$, written as

$$q_m\left(\theta \mid \widetilde{\theta}_0, \mathbf{y}_m\right) \propto \theta^{y + \frac{\alpha}{c} - 1}\left(1 - \theta\right)^{m - y + \frac{\beta}{c} - 1}.$$

Let us write the distance between the prior distribution of interest and the posterior as $\delta\left(m, \overline{\theta}, p, q_0\right)$. In this simple Bayesian model, a closed-form solution for the distance is available as the following (see Appendix 3 for full algebraic details).

$$\delta\left(m, \overline{\theta}, p, q_0\right) = \left[\frac{\alpha - 1}{\overline{\theta}^2} + \frac{\beta - 1}{\left(1 - \overline{\theta}\right)^2}\right] - \left[\frac{\overline{y} + \frac{\alpha}{c} - 1}{\overline{\theta}^2} + \frac{m - \overline{y} + \frac{\beta}{c} - 1}{\left(1 - \overline{\theta}\right)^2}\right].$$

If we are interested in the information contained in the prior $\theta \sim Beta(3,7)$ (i.e., $\alpha = 3$ and $\beta = 7$), then $\overline{\theta} = 3/(3 + 7) = .3$, and $\overline{y} = .3m$. When c is set to be very large constant:

$$\delta\left(m, \overline{\theta}, p, q_0\right) = \left[\frac{2}{.3^2} + \frac{6}{(.7)^2}\right] - \left[\frac{.3m + \frac{3}{c} - 1}{.3^2} + \frac{m - .3m + \frac{7}{c} - 1}{(.7)^2}\right] \approx \left[\frac{2}{.3^2} + \frac{6}{(.7)^2}\right] - \left[\frac{.3m - 1}{.3^2} + \frac{.7m - 1}{(.7)^2}\right].$$

Plotting the function $\delta\left(m, \overline{\theta}, p, q_0\right)$, for increasing values of m, we can see that the minimizer of $\delta\left(m, \overline{\theta}, p, q_0\right)$ is $m = 10$. The plot is a reproduction of Fig. 1 in the paper by Morita et al. [4].

2.2. Software architecture

The following code can be used to install the `R` package BayesESS.

```
R> install.packages (“BayesESS”)
```

We provide an accompanying interactive web-based application `Bayesian Effective Sample Size Calculator` to allow individuals without knowledge of the `R` programming language to easily calculate the ESS (Fig. 2). The application and its usage manual can be accessed from https://biostatistics.mdanderson.org/shinyapps/BayesESS/.

2.3. Software functionalities

The `R` package `BayesESS` can be used for determining the ESS for trivial cases where the closed-form solution exists (e.g., conjugate models such as beta-binomial, gamma-exponential, gamma-Poisson, dirichlet-multinomial, normal–normal model), and also for non-trivial cases (e.g., linear regression, logistic regression, or time-to-event model) when the ESS is determined numerically using a technique proposed by Morita et al. [4] (see Section 2.1.2; further details of the approximation strategy can be found in Appendix 5).

3. Illustrative examples

3.1. Finding the ESS in conjugate models

Below, we illustrate the use of `BayesESS R` package in identifying the ESS for the beta-binomial example discussed in the previous section. Option model=‘betaBin’ is specified for the beta-binomial model, with prior information detailed under prior. In this trivial example, the ESS is 10 as we have identified in Fig. 1.

```
R> library (BayesESS)
R> ess (model=‘betaBin’, prior=c (‘beta’, 3,7))

ESS was calculated for a beta-binomial model
ESS for the beta (alpha, beta) prior is: 10.
```

The function ess is used universally within the `R` package `BayesESS` to compute ESS for various Bayesian models. Options model and prior are used to specify the model details. Option details are available in the package manual. Further examples for using `BayesESS` in computing the ESS for conjugate Bayesian models are available in Appendix 2.

3.2. Finding ESS in linear regression model

A closed form solution is not available for many commonly used statistical models, including the linear regression (see model details in Appendix 6). The BayesESS can be used to determine the ESS for such cases, using a simulation-based numerical approximation as described by Morita et al. [4] (see details of the approximation strategy in Appendix 5). Suppose that we have a random sample of size $m, y_1, \ldots, y_m$, from a normal distribution $Y_i \mid X_i, \theta \sim N(\mu_i, 1/\tau)$, where $\mu_i = \alpha + \beta(X_i - \bar{X})$, and τ denotes a precision parameter.

$$p(y_i \mid X_i, \theta) = \left(\frac{\tau}{2\pi}\right)^{1/2} exp\left(-\frac{\tau}{2}(y_i - \mu_i)^2\right)$$
$$= \left(\frac{\tau}{2\pi}\right)^{1/2} exp\left(-\frac{\tau}{2}(y_i - (\alpha + \beta(X_i - \bar{X})))^2\right), i = 1, \ldots, n,$$

The likelihood for a random sample of size m can be written as:

$$f_m(\mathbf{y} \mid \mathbf{X}, \theta) = \left(\frac{\tau}{2\pi}\right)^{m/2} exp\left(-\frac{\tau}{2}\sum_{I=1}^{n}(y_i - (\alpha + \beta(X_i - \overline{X})))^2\right),$$

where the parameter vector $\boldsymbol{\theta}$ contains three parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\alpha, \beta, \tau)$. A prior distribution for $\boldsymbol{\theta}$ can be specified as

$$p\left(\boldsymbol{\theta} \mid \widetilde{\boldsymbol{\theta}}\right) = p_1\left(\theta_1, \theta_2 \mid \widetilde{\theta}_1, \widetilde{\theta}_2\right)p_2\left(\theta_3 \mid \widetilde{\theta}_3\right) = N\left(\theta_1 \mid \widetilde{\mu}_\alpha, \widetilde{\sigma_\alpha^2}\right) \cdot N\left(\theta_2 \mid \widetilde{\mu_\beta}, \widetilde{\sigma_\beta^2}\right) \cdot Gamma\left(\theta_3 \widetilde{a}, \widetilde{b}\right)$$

where $\widetilde{\theta}_1$ (i.e., $\widetilde{\mu}_\alpha, \widetilde{\sigma_\alpha^2}$), $\widetilde{\theta}_2$ (i.e., $\widetilde{\mu}_\beta, \widetilde{\sigma_\beta^2}$) and $\widetilde{\theta}_3$ (i.e., $\widetilde{a}, \widetilde{b}$) represent hyperparameters for θ_1, θ_2 and θ_3.

Below, we illustrate the use of `BayesESS` in identifying the ESS for a linear regression model with one covariate. Option model=‘linreg’ is specified for the linear regression model, with prior information detailed under prior. Option label is used to provide labels for the output, ncov to specify the number of covariates in the linear regression, m=50 to specify a positive integer for the maximum value in which ESS is searched, n=1000 to specify the number of simulations for numerical approximation. Further input details are discussed in the package manual.

```
library(BayesESS)
# Linear regression model with one covariate
# Priors specified as:
# beta0 ~ N(0,1), beta1 ~ N(0,1), tau ~ Gamma(1,1)
> ess(model=‘linreg’,label=c(‘beta0’,’beta1’,
‘tau’),
+ prior=list(c(‘norm’,0,1),c(‘norm’,0,1),
c(‘gamma’,1,1)),
+ ncov=1,m=50,nsim=1000,svec1=c(0,1,0),
svec2=c(0,0,1))

ESS was calculated for a linear regression model

ESSsubvector1: ESS for the first sub-vector (beta1)
ESSsubvector1: ESS for the second sub-vector (tau)

$ESSsubvec1
[1] 3.056997

$ESSsubvec2
[1] 1.9998
```

4. Impact

Bayesian inference has received increased attention from scientists seeking to incorporate prior knowledge into their modeling framework. While some simple Bayesian models (e.g., conjugate models such as beta-binomial, gamma-exponential, normal–normal models as described in later sections) offer a simple, closed-form quantification for the impact of a parametric prior [1], many commonly used statistical models such as linear or logistic regression do not extend a closed-form expression for the ESS [4], and software to evaluate the impact of prior knowledge to such modeling framework has been lacking.

We present `BayesESS`, a software for quantifying the impact of parametric priors in Bayesian analysis. The `R` package `BayesESS` also has an accompanying interactive web-application, which allows the user to further explore the impact of parametric priors through visualization. The web-application also works in place of the `R` package for applied scientists unfamiliar with the `R` programming language.

5. Conclusions

The package `BayesESS` is a comprehensive, open-source software that can be used for quantifying the impact of parametric priors in Bayesian analysis. Unique contributions of our software include: (i) capability to determine the ESS for both trivial and non-trivial cases requiring numerical estimation, (ii) accompanying web-based application for determining ESS, and (iii) additional functions to perform ESS estimation for some of the models widely used in clinical trials (see Appendices 1, 5 and 7). For efficiency, we augmented some of the key functions in the `BayesESS` package with C++ language.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

JJL's research was supported in part by the grants CA016672 and CA221703 from the National Cancer Institute, United States.

Data availability

No data was used for the research described in the article.

References

[1]. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian data analysis. Chapman and Hall/CRC; 2014.

[2]. Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012;31:2955–72. [PubMed: 22711340]

[3]. Berry D Bayesian clinical trials. Nat Rev 2006;5:27–36.

[4]. Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. Biometrics 2008;64(2):595–602. [PubMed: 17764481]

Author Manuscript
Author Manuscript
Author Manuscript

[5]. Morita S, Thall PF, Müller P. Evaluating the impact of prior assumptions in Bayesian biostatistics. Stat Biosci 2010;2:1–17. [PubMed: 20668651]

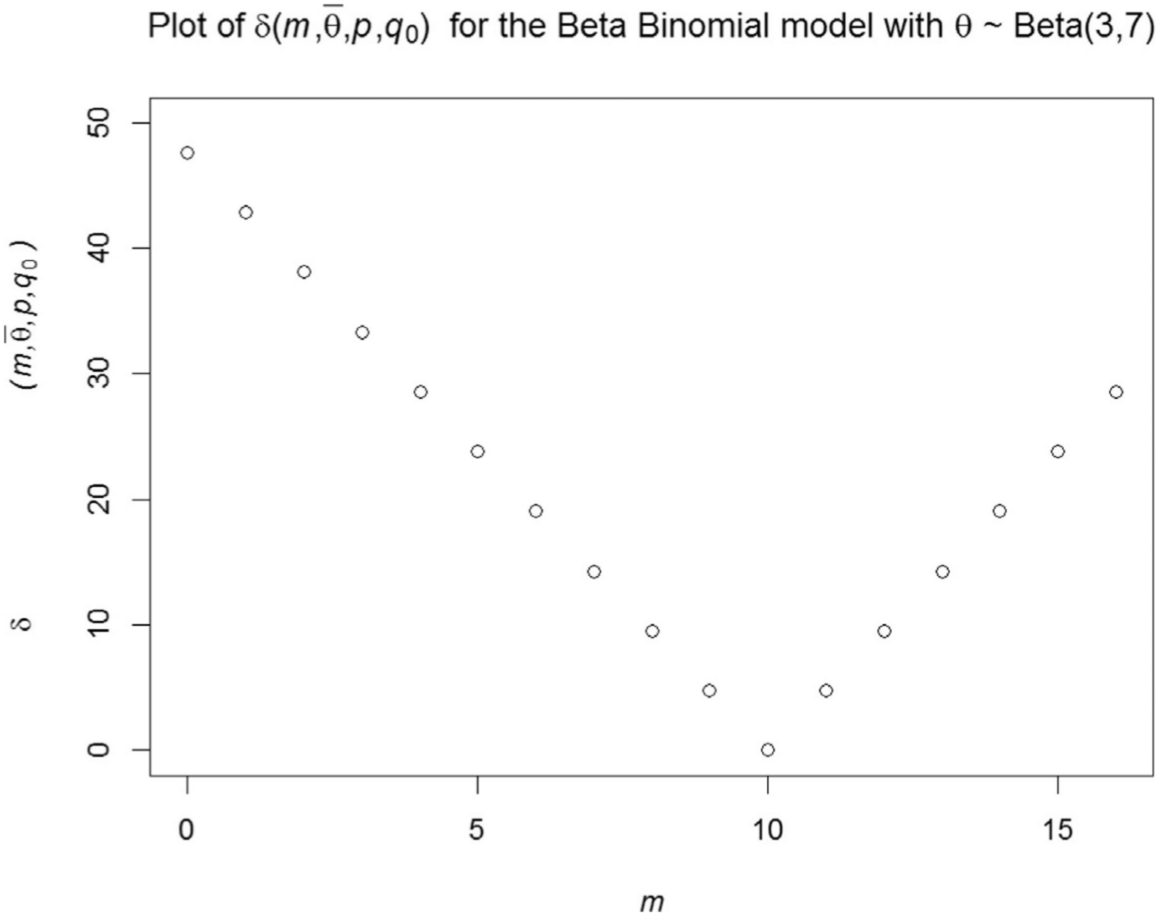

Fig. 1.
Plot of function $\delta\left(m, \overline{\theta}, p, q_0\right)$.

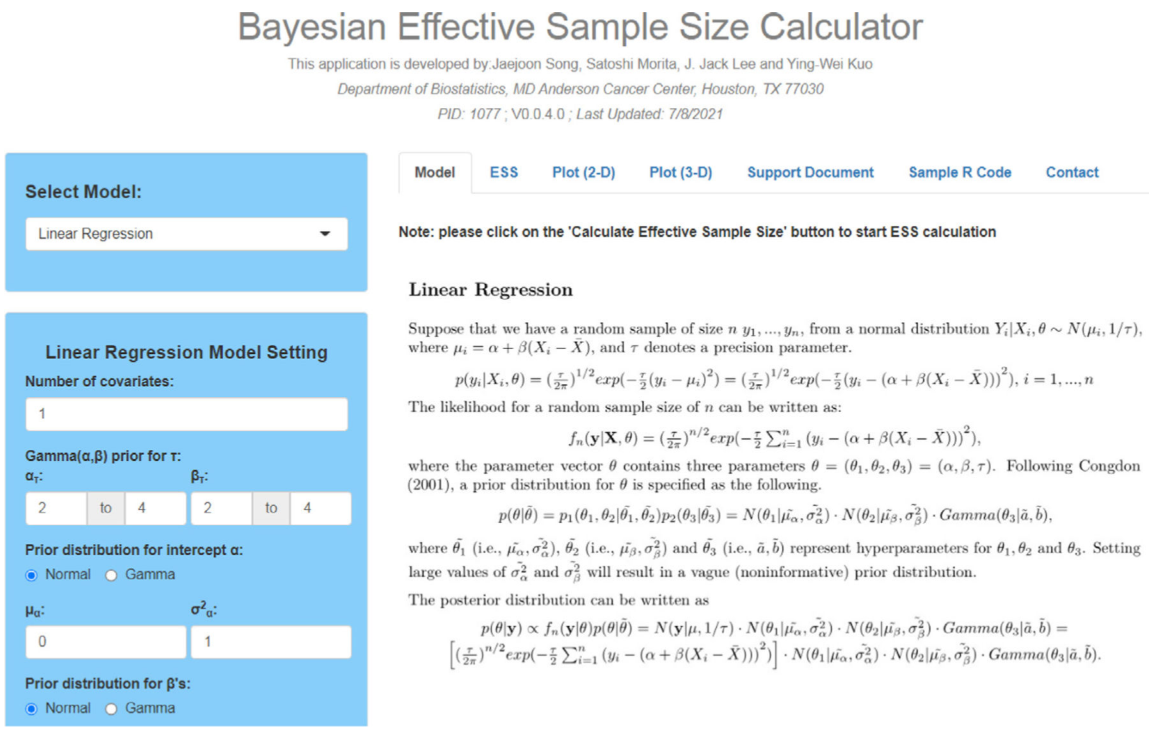

Fig. 2.

ESS calculation using the web-based application.