THE UNIVERSITY OF TEXAS

MD Anderson Cancer Center

Making Cancer History®

The University of Texas MD Anderson Cancer Center
Division of Quantitative Sciences
Department of Biostatistics

Beta-Binomial Distribution Demo

User's Guide Version 1.0

J. Jack Lee, Ying-Wei Kuo and Clift Norris
02/06/2015

Contents

Overview

The beta-binomial distribution is frequently used in Bayesian statistics to model the number of successes in *n* trials. The main purpose of the software is to illustrate how a prior distribution is updated to become a posterior distribution after observing the data via the relationship of the beta-binomial distribution.

This demo program has two options for displaying the Bayesian process, which are organized into the two tab pages labeled "Step-by-Step Learning" and "Trial Simulation".

This program is distributed at no cost to the user. However, redistribution of this program is not permitted. Each person should obtain a copy directly from The University of Texas MD Anderson Cancer Center at http://biostatistics.mdanderson.org/SoftwareDownload/.

This allows us to keep a record of who may be using the software and allows us to notify all users when program enhancements become available.

System Requirements

- Windows 7 SP1
- Microsoft .NET Framework version 4.5 (full framework, x86 and x64)
- Windows Installer 4.5
- Minimum screen resolution 1024x768

If any of the required software components is missing, the installation procedure will install them.

Disclaimer

We provide absolutely no warranty of any kind, expressed or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the program lies with the user.

Should this program prove defective, the user assumes the cost of all necessary servicing, repair, or correction.

In no event shall The University of Texas or any of its component institutions, including MD Anderson Cancer Center, be liable for damages, including any lost profits, lost monies, or other special, incidental or consequential damages arising out of the use of or inability to use (including but not limited to loss of data or its analysis being rendered inaccurate or losses sustained by third parties) the program.

1   Introduction

The posterior distribution of the parameter summarizes what is known about the parameter by combining the prior distribution and the observed data.

The **Beta-Binomial distribution** forms the distribution of the number of successes, assuming the probability of success, i.e., the parameter of the binomial distribution at each trial is not fixed but random. We assume that the probability of success follows a beta distribution. Given the probability of success, the number of successes after n trials follows a binomial distribution. Note that the beta distribution is a conjugate prior for the parameter of the binomial distribution. In this case, the likelihood function is binomial, and a beta prior distribution yields a beta posterior distribution. For example, assuming the likelihood follows a *binomial(n, θ)* distribution, where n is known and θ is the parameter of interest, and the number of successes x is an integer between 0 and n, then:

Prior of θ	*Beta(a,b)*
Likelihood	*binomial(n, θ)*
Posterior of θ	*Beta(a+x, b+n– x)*
Posterior Mean	*(a+x)/(a+b+n-x)*

A variable with a beta-binomial distribution is distributed as a binomial distribution with parameter p, where p is distributed as a beta distribution with parameters a and b. For n trials, the probability density function of x follows:

$$p(x) = \frac{beta(x + a, n - x + b)\binom{n}{x}}{Beta(a,b)}$$

where *beta(a, b)* is a beta function and $\binom{n}{x}$ is a binomial coefficient.

For more details, see the information at the following URLs:
http://en.wikipedia.org/wiki/Conjugate_prior_distribution
http://mathworld.wolfram.com/BetaBinomialDistribution.html

2   Step-by-Step Learning

The "**Step-by-Step Learning**" tab is constructed to show the prior and posterior distributions of the success rate of the binomial distribution. Assuming the number of successes follows a binomial distribution and the prior distribution of success follows a beta distribution, the posterior distribution also follows a beta distribution. To learn how the posterior distribution is formed, input all parameter values into the appropriate edit boxes. The output of the calculation will be displayed in the panels below the input parameters, which are initially empty.
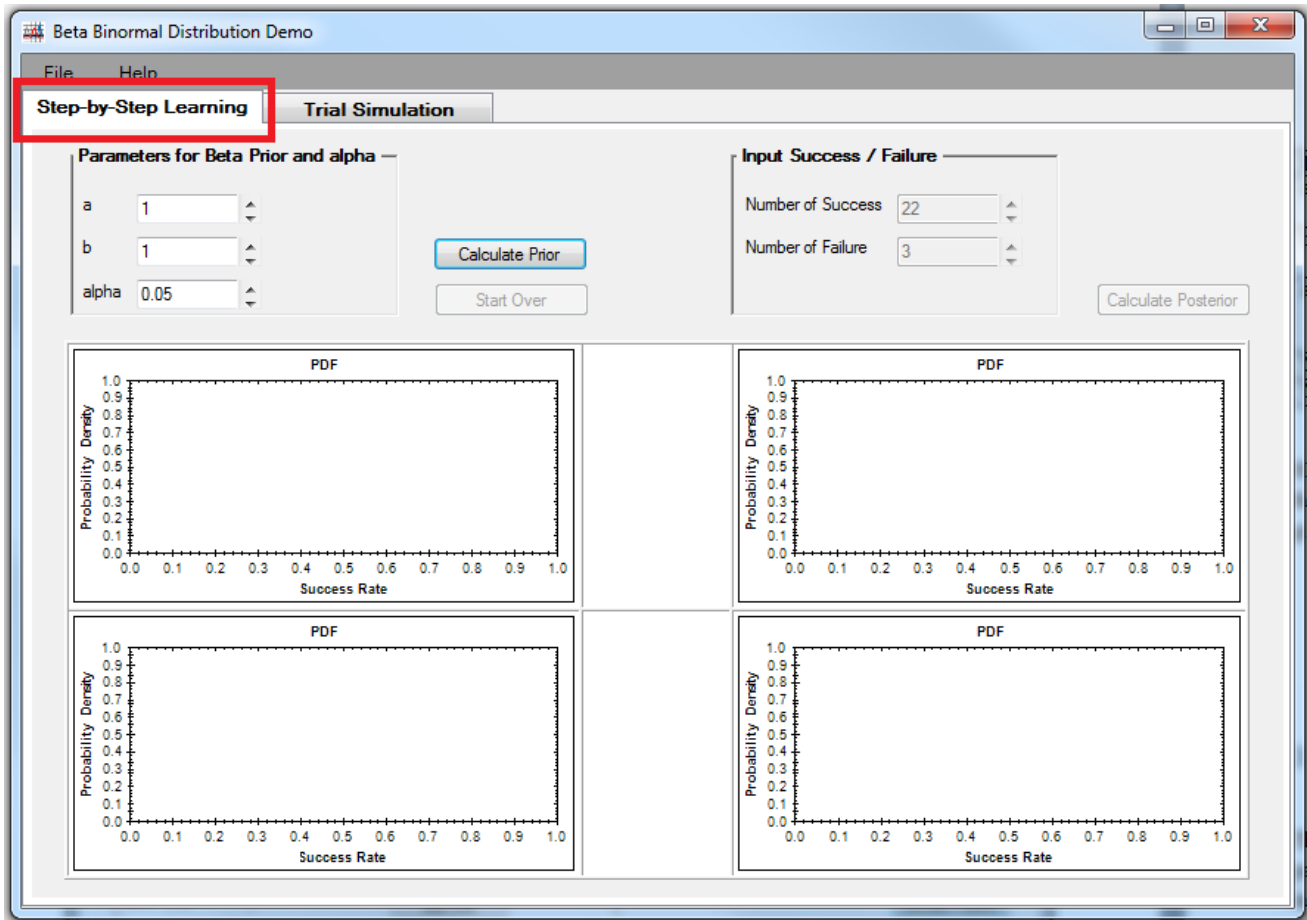

2.1   Parameters for Beta Prior and alpha

Two parameters of the conjugate beta prior in this learning process are specified in the **a** and **b** fields. Note: *Beta(a,b)* denotes the beta distribution, where a > 0 and b > 0. A credible interval for the probability of success is computed as an interval in the domain of a posterior probability distribution for interval estimation. The level of confidence is (1 – **alpha**) *100% *where 0 ≤ alpha ≤ 1.* The highest probability method is used to compute the credible interval.

2.2   Input Success / Failure

The number of successes and the number of failures in the trial are recorded in the **Number of Success** and **Number of Failure** fields. Both parameters are non-negative integers.

2.3 Instructions:

1. Input parameters for beta prior and alpha.
2. Click the "Calculate Prior" button.
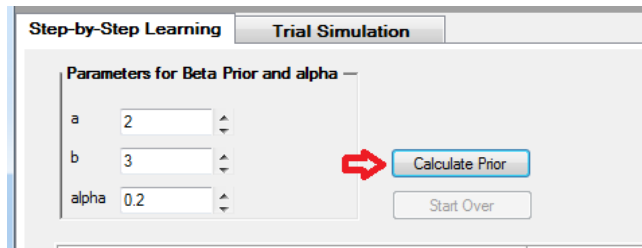
After the calculation is done, the probability density function of the prior will be shown in **Figure 1**. Its credible interval is displayed on the plot panel. The prior distribution, its mean, and the (1– alpha) *100% credible interval are displayed in the figure title. The "Calculate Prior" button and parameters for the beta prior and alpha are all disabled at this point. *Note: Click the "Start Over" button to restart the calculation if you want to change the prior values.*
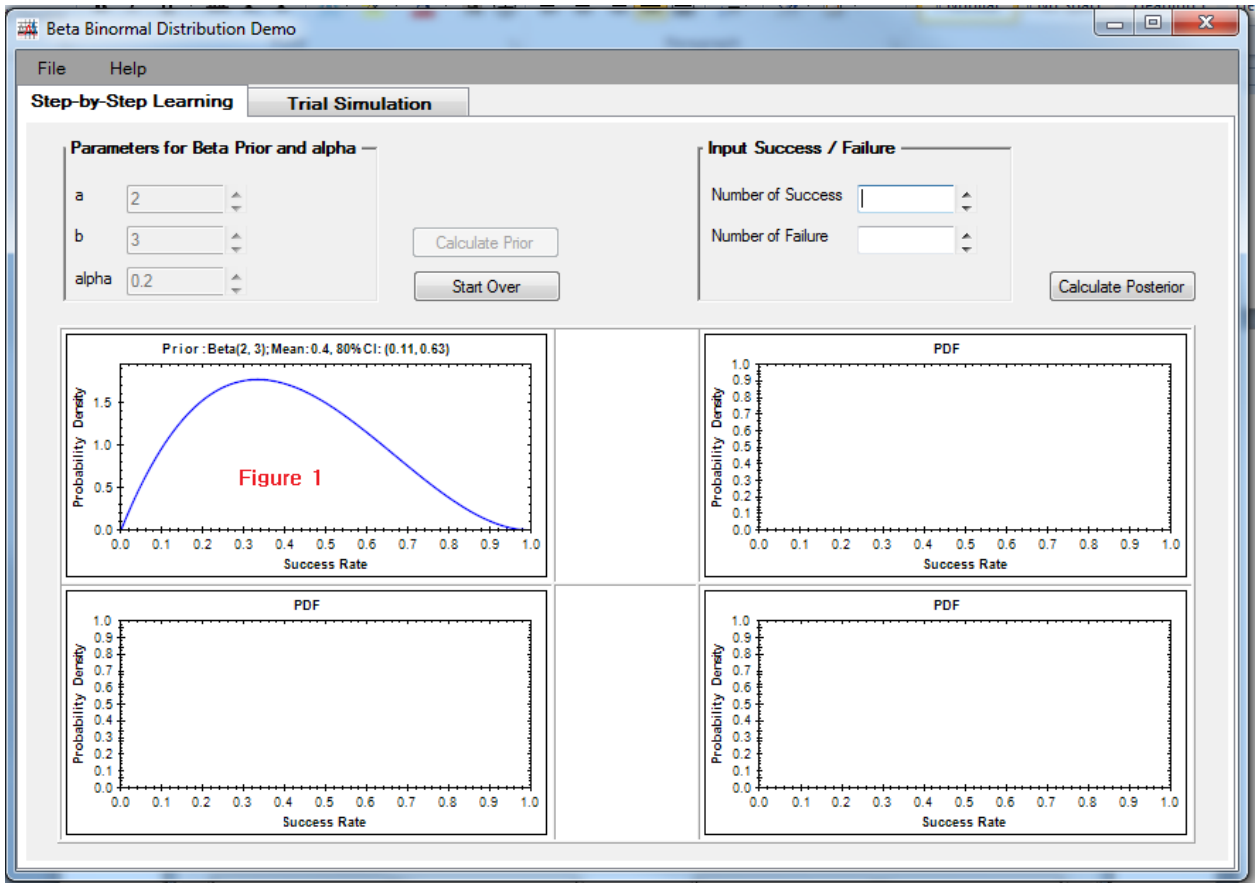

3. Input the number of successes and number of failures.
4. Click the "Calculate Posterior" button.
   After the calculation is done, the probability density function of the posterior will be displayed in **Figure 2**. The red curve represents the posterior distribution and the blue curve represents the prior distribution, which is the same as the blue curve in **Figure 1**. *Note: Again, the user may click the "Start Over" button at any time to restart the calculation.*

6
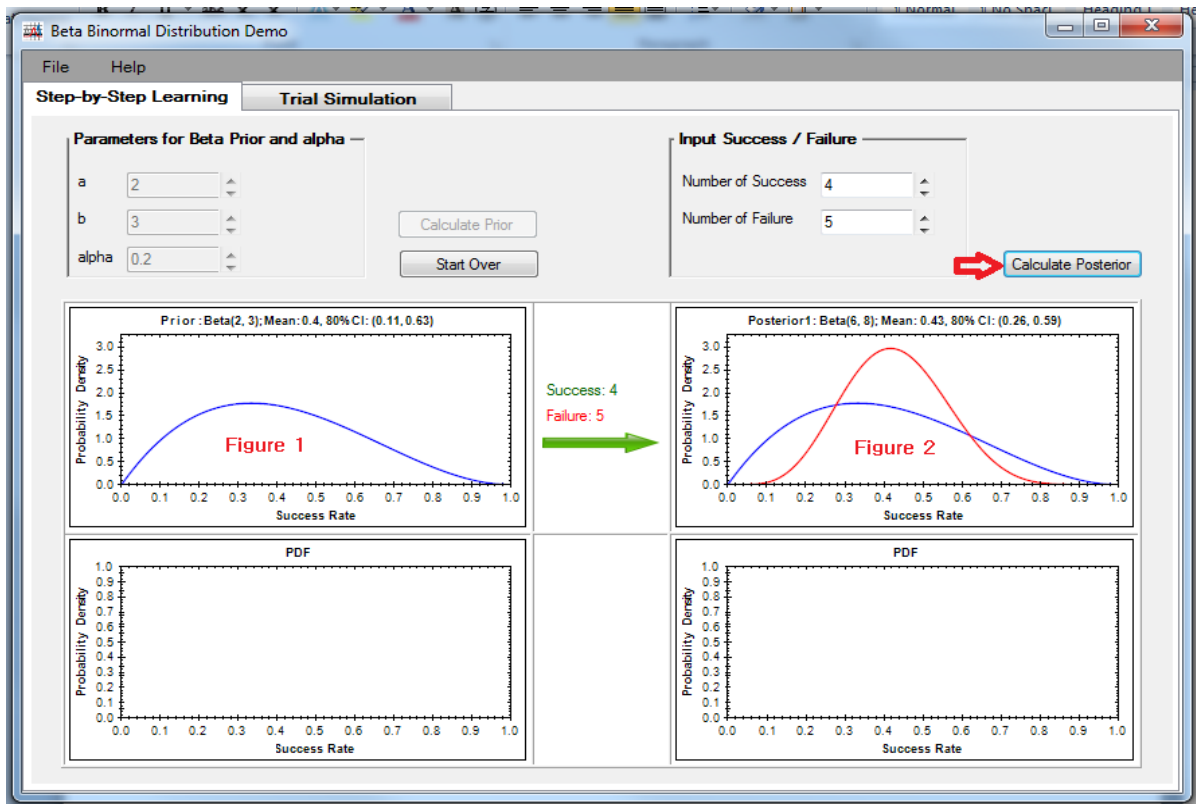
5. Repeat step 3 and step 4 to continue calculating another set of posterior distributions. The latest posterior results will always be displayed as a red curve, as shown in **Figure 4**. The blue curve in **Figure 3** (and **Figure 4**) represents the prior of the posterior in **Figure 4,** and the prior in **Figure 3** was the posterior in **Figure 2**.
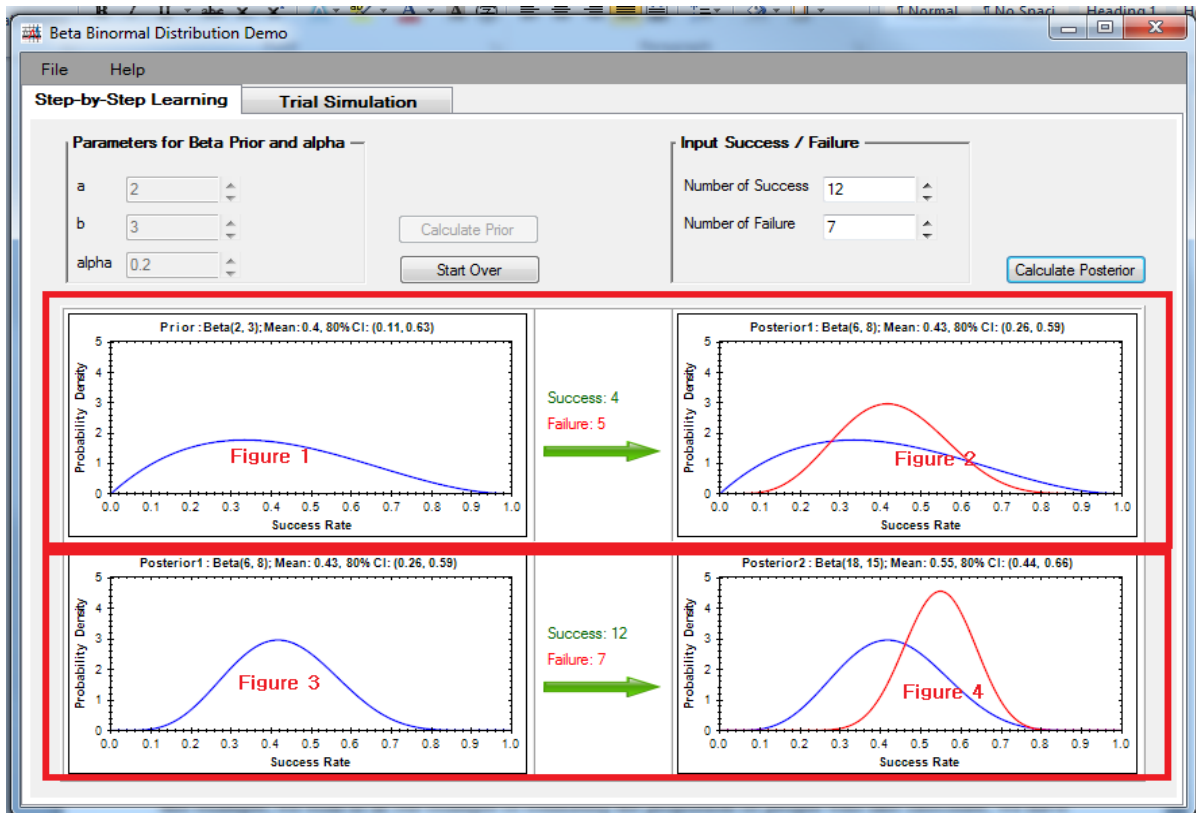
3 Trial Simulation

The "**Trial Simulation**" tab is used to simulate a series of binomial experiments, such as observing the outcome of success or response in clinical trials. This simulation contains three types of input parameters: Parameters for the Beta Prior and alpha, Design Parameters and Simulation Setting.
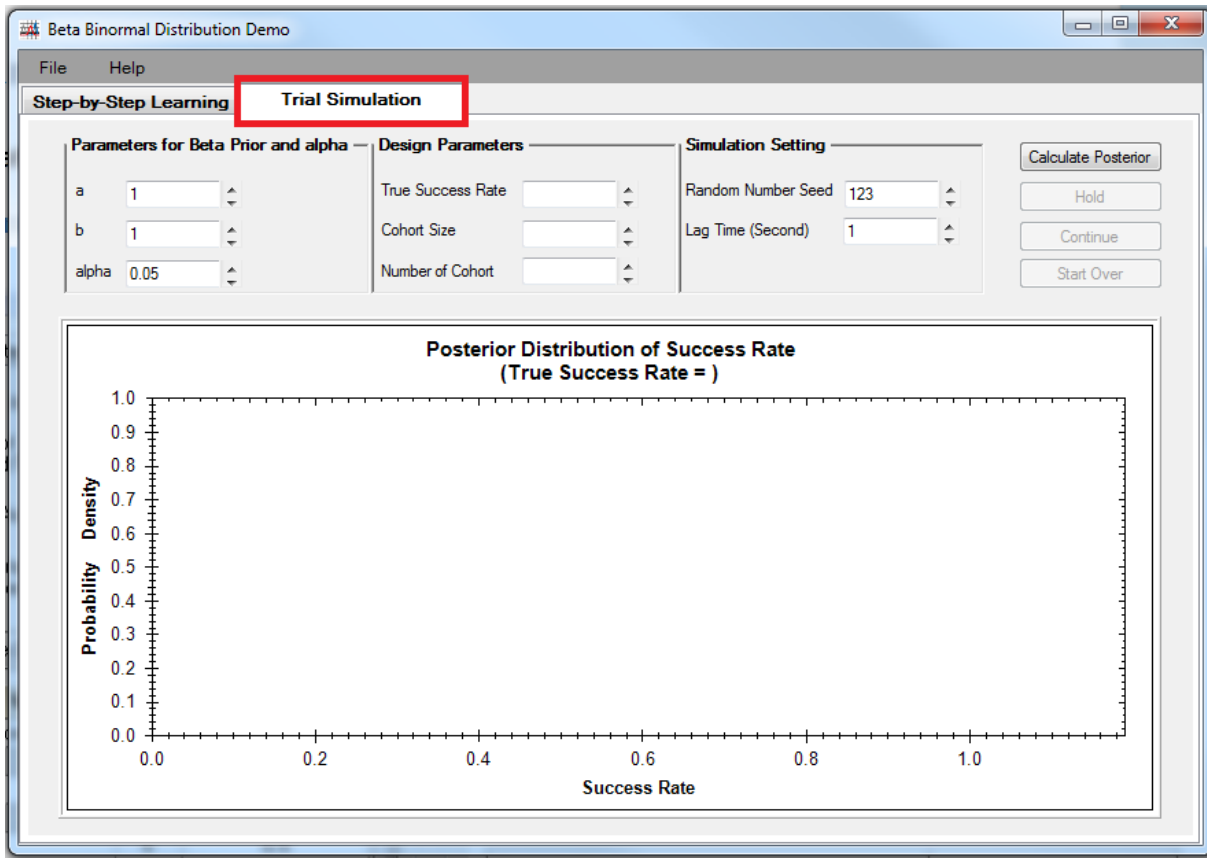
3.1 Parameters for Beta Prior and alpha

Two parameters of the conjugate beta prior in this learning process are specified in the **a** and **b** fields. Note: *Beta(a,b)* denotes the beta distribution, where a > 0 and b > 0. A credible interval for the probability of success is computed as an interval in the domain of a posterior probability distribution for interval estimation. The level of confidence is (1 – **alpha**) *100% *where 0 ≤ alpha ≤ 1.* The highest probability method is used to compute the credible interval.

3.2 Design Parameters

The **True Success Rate** field is the probability of success in a trial. **Cohort Size** represents the number of patients enrolled in a trial as a group at one time and **Number of Cohort** is the number of groups in a trial.

3.3 Simulation Setting

The seed is specified in the **Random Number Seed** field, which allows users to reproduce results by using the same seed number. The seed number is the starting point used in the generation of a sequence of random numbers;

therefore, users may obtain the same results given the same seed number. The **Lag Time (Second)** field controls the delay between the displays of output for each cohort.

3.4 Instructions:

1. Enter all the required input parameters and click the "Calculate Posterior" button.
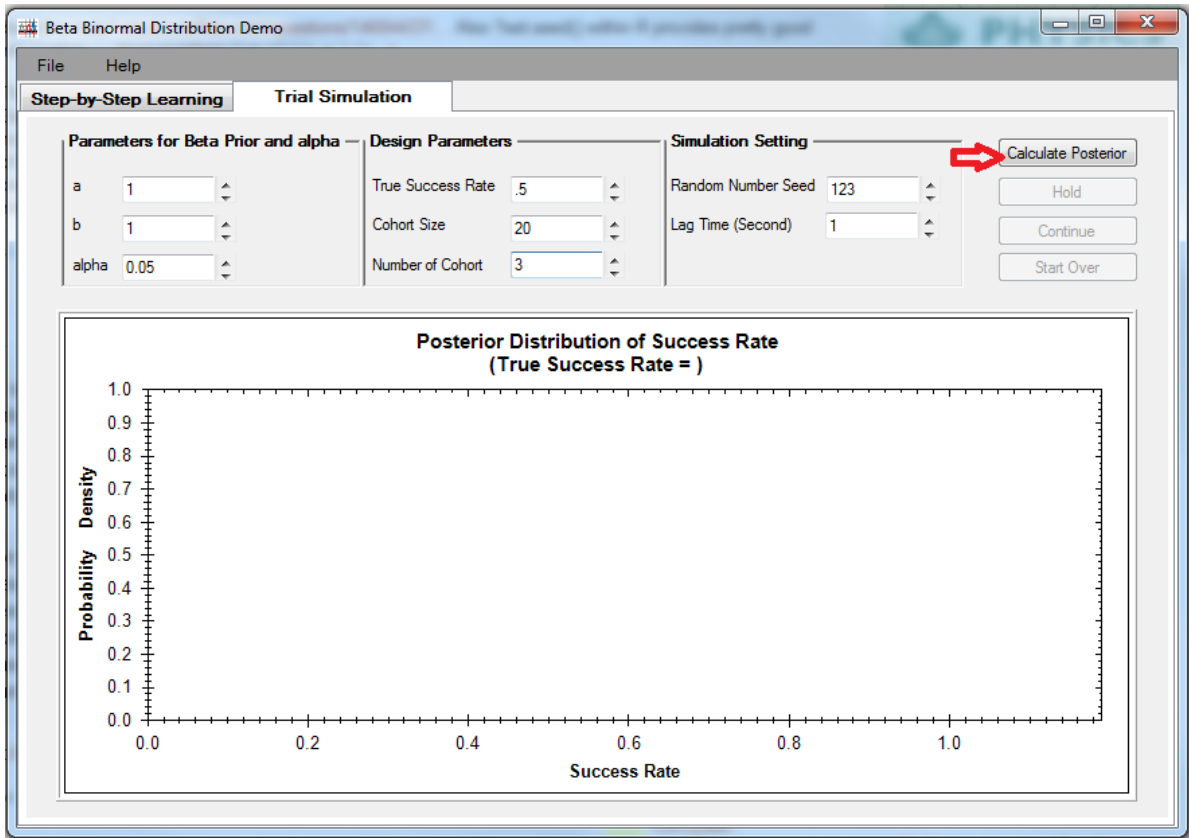
2. The result for each cohort will be displayed in the output window. The following information will be contained in the legend of the output figure:

    *   Cohort number
    *   Number of successes for the specified cohort
    *   Number of failures for the specified cohort
    *   Cumulative number of successes
    *   Cumulative number of failures
    *   Posterior mean
    *   Credible interval of the success rate

3. All the input parameters and the "Calculate Posterior" button will be disabled at this point. Users may click the "Hold" button at any time to freeze the output screen and then click the "Resume" button when ready to continue the simulation. Users may also click the "Start Over" button to restart the simulation. The red curve shows the latest posterior distribution. The blue curve shows the immediate prior distribution. All the earlier distributions are shown in gray. When the simulation is completed, users may edit the **Number of Cohort** and then click the "Continue" button to continue the simulation. Or users may click the "Start Over" button to begin a new simulation.

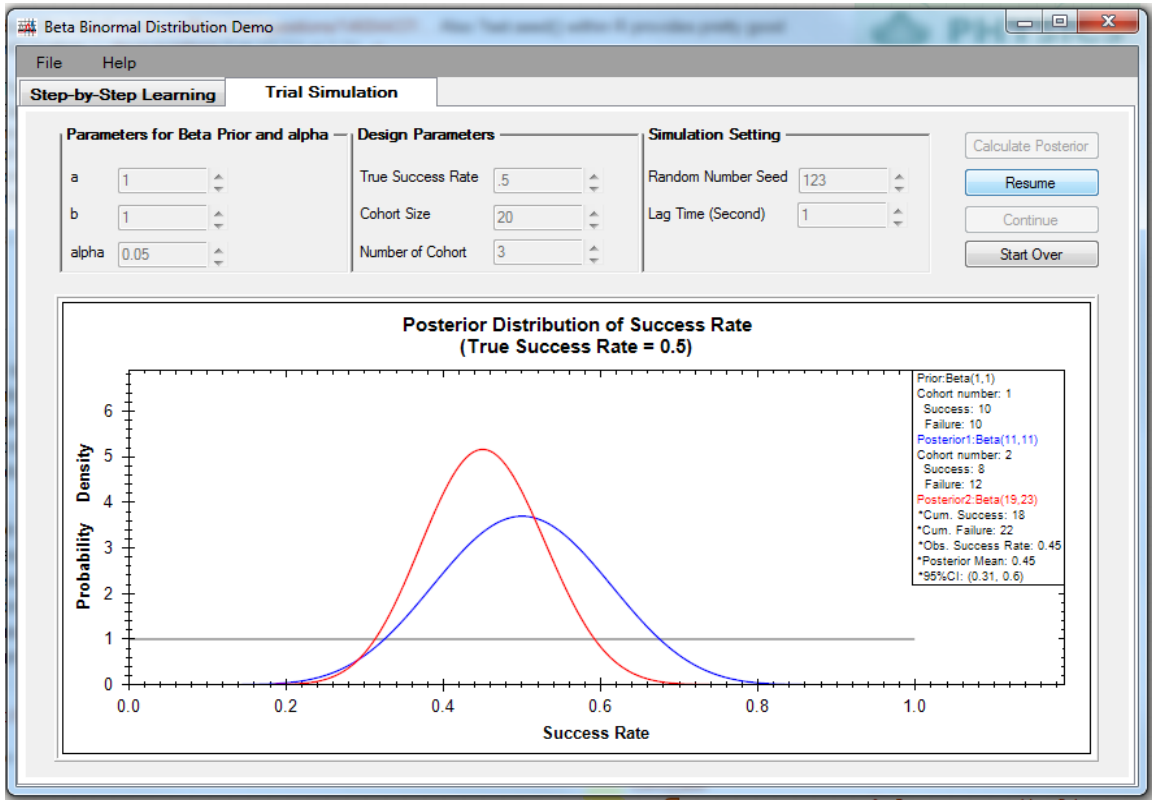   An example of a sequence of the output is shown below:
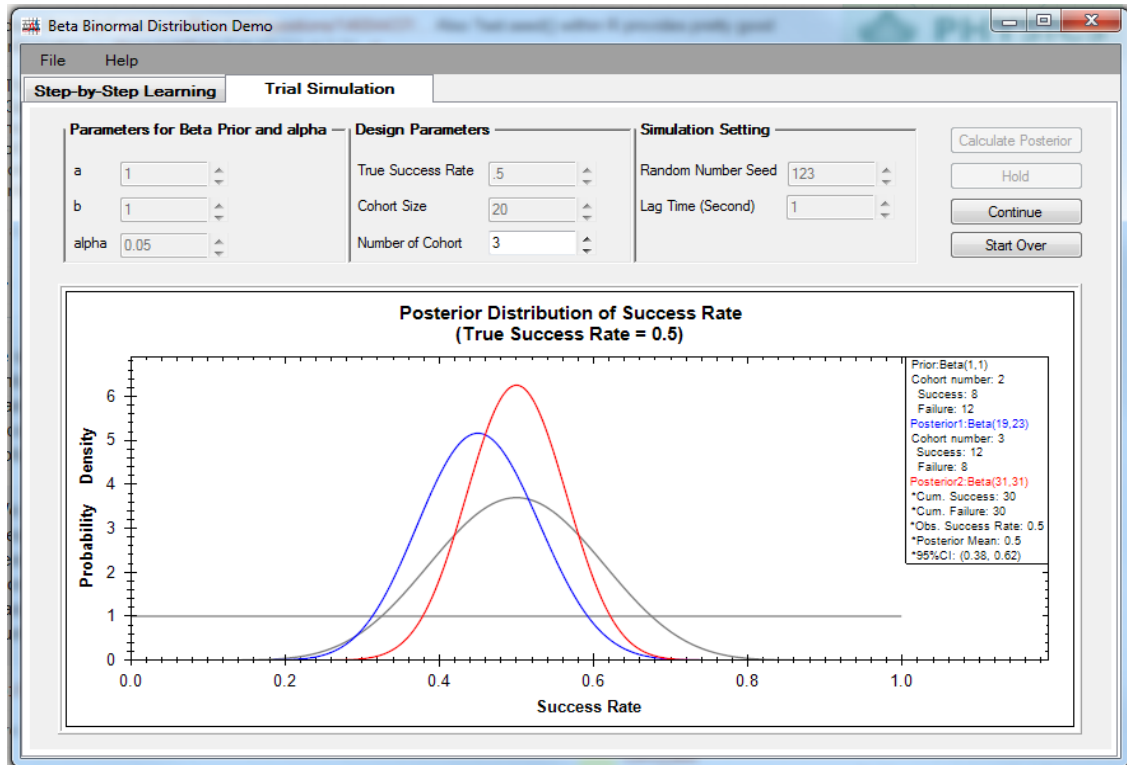
Prior: Beta(1,1)
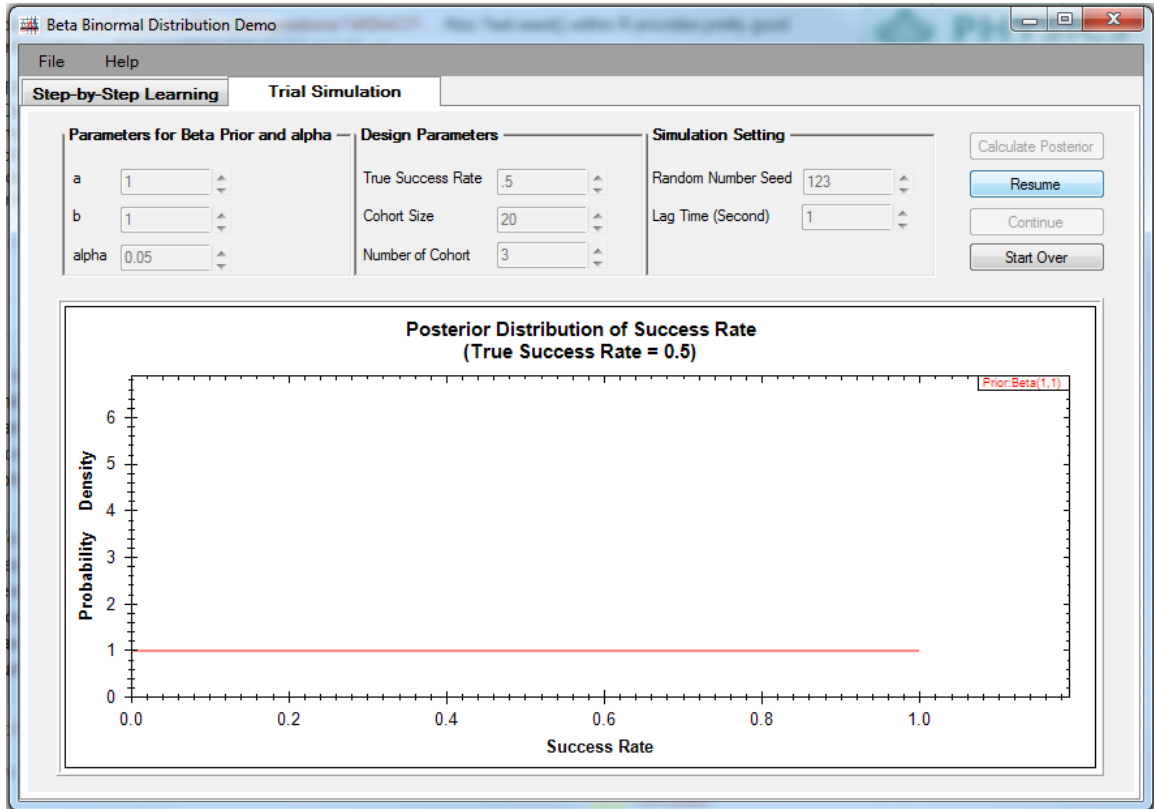
Cohort Number:1
Number of Success: 10; Number of Failure: 10; Posterior: Beta(11,11)

10
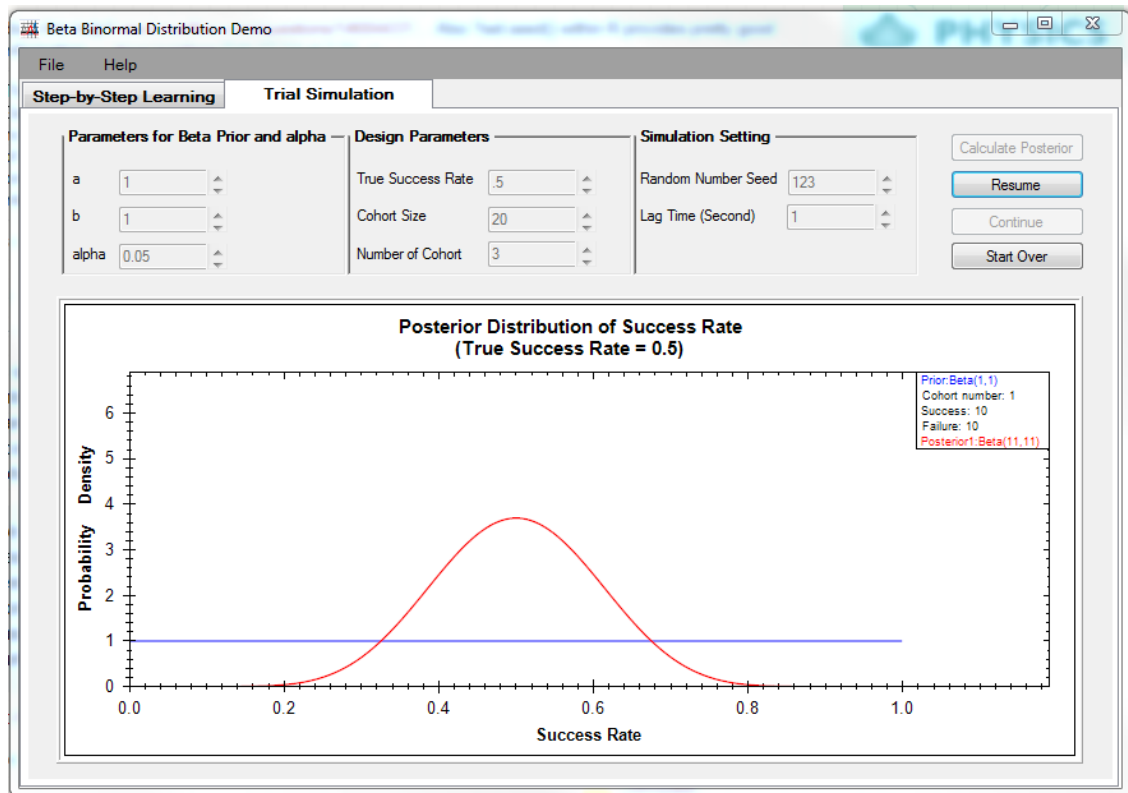
Cohort Number: 2
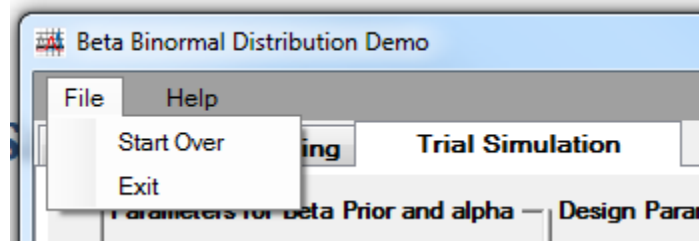## Number of Success: 8; Number of Failure: 12; Posterior: Beta(19,23)

Cohort Number: 3
Number of Success: 12; Number of Failure: 8; Posterior: Beta(31, 31)

4 File Menu and Help Menu

4.1 File Menu

- The calculation can be terminated and restarted via *File → Start Over*
- The demo program can be closed via *File → Exit*

4.2 Help Menu

- The User's Guide (PDF version) of this program can be viewed via *Help → User's Guide*
- The program's license information can be found via *Help → View License*
- Other information, including the program's version number, can be found via *Help → About*