

Abstract

In most statistical analyses the data is collected using a fixed sample size and is then analyzed using the complete dataset. In contrast, sequential sampling takes one sample at a time and analysis is performed at each step. Sampling in this way continues until sufficient information in the cumulative sample is obtained to make an appropriate decision using statistical inference. Thus by using sequential sampling instead of fixed samples, one can draw conclusions during the data collection and often reach a decision more quickly. Sequential inference methods have the potential to greatly reduce the time and costs associated with decision making processes. As a result, they are applied to a variety of different fields such as Entomology, where the average pest density number is the parameter of interest, and to the study of clinical trials.

The mathematics associated with designing and utilizing sequential procedures can be complex for practitioners and as such can dissuade users from availing themselves of the advantages they offer. We propose to develop a software tool that facilitates an easier use of the methods, and enables a convenient platform for carrying out sequential inference. The tool implements the appropriate inference procedures associated with each design within an innovative and scalable software architecture. It also hides the complex mathematics of the designs, and provides a user-friendly interface to design a sequential sampling scheme, understand its important statistical properties, and implement the scheme on real data. This tool not only allows the user to analyze the data in real time, but also saves the data and all other text and graphical outputs for use in subsequent analyses. The software design and architecture allows for the scalability and portability across the platform of operating systems.

Background

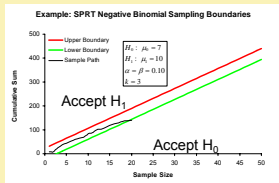
The Basic SPRT:

At each stage of the test, an observation is taken from the population under consideration and the likelihood ratio:

$$\lambda_n = \prod_{i=1}^n \frac{f(x_i; \theta_1)}{f(x_i; \theta_0)}$$

is constructed. Based on the values of the likelihood ratio, one of the following decisions is made: a) Accept H_0 if $\lambda_n \leq B$
b) Accept H_1 if $\lambda_n \geq A$
c) Continue sampling otherwise

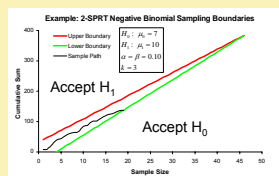
Wald's approximation is used to choose $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$ in order to achieve: $\Pr(\text{Reject } H_0 | H_0) = \alpha$ and $\Pr(\text{Reject } H_1 | H_1) = \beta$



The 2-SPRT

$$\lambda_{1n} = \prod_{i=1}^n \frac{f(x_i; \theta_0)}{f(x_i; \theta')}, \quad \lambda_{2n} = \prod_{i=1}^n \frac{f(x_i; \theta')}{f(x_i; \theta'')}$$

are constructed. Based on the values of the likelihood ratio, one of the following decisions is made: a) Accept H_0 if $\lambda_{1n} \leq B$
b) Accept H_1 if $\lambda_{2n} \geq A$
c) Continue sampling otherwise



Advanced Sequential Sampling Methods and Their Implementation Within a Portable Computing Environment

Arunava Chakravarty¹, Payal Shah¹, Carlos Rendon²
Daniel R. Jeske¹, Pengyue J. Lin³, Robert Luck⁴

¹Department of Statistics, College of Natural and Agricultural Sciences, University of California, Riverside, CA 92521
²Department of Computer Science, College of Engineering, University of California, Riverside, CA 92521
³College of Humanities, Arts and Social Sciences, University of California, Riverside, CA 92521
⁴Department of Entomology, University of California, Riverside, CA 92521

Sequential Bayes Procedure

We test $H_0: \mu = \mu_0$ vs $\mu = \mu_1$ where $x_i \sim f(\mu; \delta)$
 δ is a nuisance parameter
We assume a discrete prior $g(\mu, \delta)$ for (μ, δ) , $i \in \{0,1\}$, $j \in \{1, \dots, n\}$

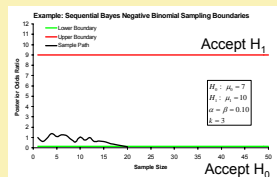
Step 1:
For the k^{th} observation x_k we get the posterior probabilities of $\mu, \delta; i \in \{0,1\}, j \in \{1, \dots, n\}$ as

$$p(\mu, \delta_j | x_1, \dots, x_k) \propto f(x_1, \dots, x_k | \mu, \delta_j) \times g(\mu, \delta_j)$$

We now compute the marginal posterior probabilities of $\theta_i, i \in \{0,1\}$ as

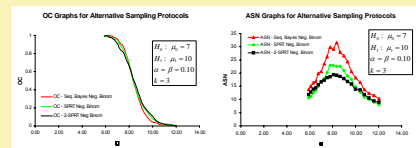
$$p(\mu_i | x_1, \dots, x_k) \propto \sum_{j=1}^n p(\mu, \delta_j | x_1, \dots, x_k)$$

Step 2:
If $p(\mu_0 | x_1, \dots, x_k) \geq \rho$ then accept H_0 else if $p(\mu_1 | x_1, \dots, x_k) \geq \rho$ then accept H_1 else $g(\mu, \delta_j) = p(\mu, \delta_j | x_1, \dots, x_k)$ and obtain another observation x_{k+1} and go to Step 1



OC and ASN Curves

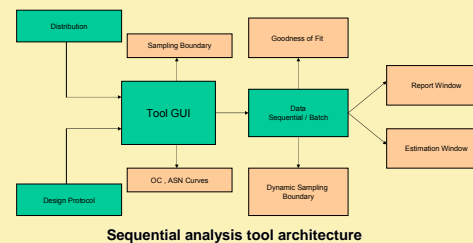
- The term 'operating characteristic function', or better known as the OC function reports the probability of acceptance.
- The term 'average sample number function', often shortened to ASN function reports the average amount of samples required in order to make a decision.



Objectives

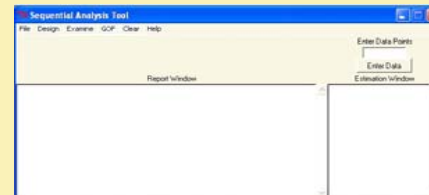
- Facilitate easier use of advanced sequential sampling methods
- Conceal the complex mathematics of the design and implementation of the methods from field practitioners
- Develop a laptop software tool to implement the appropriate inference procedures associated with each design
- Design an evolutionary and scalable software architecture
- Provide a user-friendly interface to the features of the tool

Methods

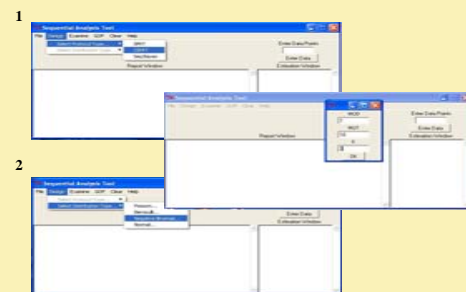


- User selects design protocol and model for the distribution of the data
- User can examine shape and position of sampling boundaries
- User can examine the OC and ASN functions for the sampling design
- User sequentially enters the observations and can view the dynamic sampling boundary and the estimates under the assumed model
- User can save graphs and reports to external files

Results

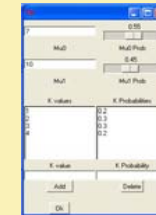


Sequential analysis tool main menu



Three step process of choosing the design protocol and model

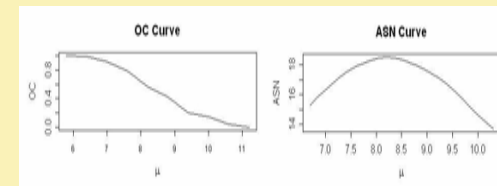
Sequential Bayes Procedure



Window to input the prior distributions of the parameters of interest for the sequential bayes procedure

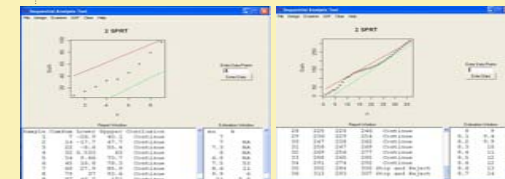
- In this parameter input window, the user is required to input both the parameter values and their prior probabilities to initiate the Sequential Bayes procedure.
- The rest of the outputs are similar to the ones for the other protocols.

OC and ASN Curves



Screen shots of the OC and ASN curves from the tool

Sequential Data Generated Outputs



Dynamic graphs show the process of entering the sequential data, a report window keeps track of all the steps taken in the procedure and an estimation window gives current estimates of the parameters at each step of the procedure

Conclusion

- This tool serves as an interface between the theory of sequential analysis and the end user practitioner.
- The tool provides the ability to compare and contrast alternative sequential sampling designs.
- The tool hides the complicated mathematical computations to make the application of sequential methods easier.
- The tool works well on laptop which facilitates its use in field environments.
- Future enhancements could be made by incorporating sequential protocols used in clinical trial applications.