Sequential Probability Ratio Test (SPRT) by R

by **Behzad FallahiFard**

behzadfalahifard@gmail.com

Supervisor: **Dr. Aliakbar Rasekhi**

Assistant Professor at the Faculty of Medical Sciences

rasekhi@modares.ac.ir

Tarbiat Modares University, Tehran, Iran

Abstract

In this project, a function called `sprt` is presented which has performed statistical testing of a simple hypothesis (H_0) versus another simple hypothesis (H_1) using the SPRT method. Also, another function called `en.sprt` is designed to estimate the number of sampling steps for accepting H_0 or accepting H_1.

Introduction

In Neyman-Pearson statistical tests, the probability of statistical errors depends on the number of observations. In other words, by changing the number of observations, the probability of statistical errors can also be altered. However, in such tests, we keep the number of observations constant and reduce the probability of statistical errors.

Neyman-Pearson tests, especially Uniform Most Powerful (UMP) tests, are obtained when we can change the probability of errors, and the sample size is fixed. But in sequential tests, some methods are used to optimize sample size based on fixed probability of type I and type II errors. These tests are based on the likelihood ratio tests.

Problem 1

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with probability density function f. Suppose that f_0 and f_1 are two possible densities for $X_i s$. f_0 and f_1 are both related to continuous distributions or discreet distributions with the same support.

Consider the problem of testing

$$H_0: f = f_0 \quad vs \quad H_1: f = f_1$$

Definition:

$$R_m(x_1, x_2, \ldots, x_m) = \frac{f_1(x_1, x_2, \ldots, x_m)}{f_0(x_1, x_2, \ldots, x_m)} = \prod_{i=1}^{m} \frac{f_1(x_i)}{f_0(x_i)}$$

Wald (1945) proposed the following:

Let $0 < B < 1 < A < \infty$:

After observing $X_1 = x_1$

Step 1:

If $R_1(x_1) \geq A$, stop sampling and accept H_0

If $R_1(x_1) \leq B$, stop sampling and accept H_1

If $A < R_1(x_1) < B$, continue to sampling and consider $X_1 = x_1$, $X_2 = x_2$ for step 2

Similarly for step m:

If $R_m(x_1, \ldots, x_m) \geq A$, stop sampling and accept H_0

If $R_m(x_1, \ldots, x_m) \leq B$, stop sampling and accept H_1

If $B < R_m(x_1, \ldots, x_m) < A$, continue to sampling for step m+1

Definition:

For two constant values A and B, the method presented before is called Sequential Probability Ratio Test (SPRT); Which is also the "best" sequential method.

In the SPRT method, we show the sample size with the symbol N, which is a random variable. In this method, it is proved that sampling ends; we do not need infinite observation. A logical approximation for A and B can be as follows

$$A \approx \frac{1 - \beta}{\alpha} \quad \& \quad B \approx \frac{\beta}{1 - \alpha}$$

Example 1

Let $S\,F\,F\,S\,S\,S\,F\,S\,\ldots$ be an independent sequence of Bernoulli distribution, $B(1, \theta)$. Consider the problem of SPR testing

$$H_0: \theta = \frac{3}{10} \quad vs \quad H_1: \theta = \frac{6}{10}$$

If the probability of type I and type II errors are 0.2 ($\alpha = \beta = 0.2$). At what step is the sampling stopped, and how is the test decision made?

Programming 1

> # B.Falahi Fard 881537

> #   sprt : Sequential Probability Ratio Test

> # Basic concepts of mathematical statistics by: Dr.A.Parsian second edition page:436

> # If 'xm' is observed sample in step m:

> # fH0 : f(x1,x2,...,xm | H0)

> # fH1 : f(x1,x2,...,xm | H1)

> # alpha: type I Error

> # beta : type II Error

> # R[m] = f(x1,x2,...,xm | H1)/f(x1,x2,...,xm | H0)--

> sprt<-function(fH0,fH1,alpha,beta){

+ A <-(1-beta)/alpha

+ B <- beta/(1-alpha)

+ f1 <-fH1

+ f0 <-fH0

+ cp1<- cumprod(f1)   # If B < R[m] <A then continue sampling.

+ cp0<- cumprod(f0)   # If R[m] > A then H0 is rejected by N=m observation.

+ m<- length(cp1)   # If R[m] < B then H0 is accepted by N=m observation.

+ R<- rep(0,m)

+ R<-cp1/cp0

+ y<- R

+ x<- rep(1:m)

+ if(R[m]>B & R[m]<A){

+ cat("\n","   B < R","[",m,"]",

```
+ "< A    ","\n","",B,"<",R[m],"<",A,"   -->",
+ " Countinue sampling.","\n","\n")}
+ else if(R[m]>A)
+ cat("\n","R","[",m,"]","> A    ","\n",R[m],">",A,"     -->",
+ " Rejcet H0.Finished","\n","\n")
+ else cat("\n","B >","R","[",m,"]","\n",B,">",R[m],"     -->",
+ " Accept H0.Finished","\n","\n")
+ if(R[m]<B | R[m]>A) {
+ plot(1,R[1],type="o",xlim=c(0,m),ylim=c(0,A+3),xlab="N",
+ ylab=expression(R[N]), main=expression('nemodare masire '(N,R[N])));
+ u<-c(0,0,m,m+100);v<-c(0,B,B,0);polygon(u,v,col='green');
+ z<-c(A,A+100,A+100,A);polygon(u,z,col='7');
+ lines(x,y,type="o",lwd=2);
+ abline(h=A,col='red');abline(h=B,col='red');
+ text(m/2,A+1.5,"nahieye radde H0",col='blue',lwd=2);
+ text(m/2,B/2,"nahieye ghabole Ho",col='blue',lwd=2);
+ }} #_____ Solve :_____
> # z=(1,0,0,1,1,1,0,1,1, ... )
> #-------------------------------------- --------------------------------------------------------Step 1:
> x<-c(1)
> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

   B <  R [ 1 ] < A

 0.25 < 2 < 4   -->  Countinue sampling.
> #-------------------------------------- --------------------------------------------------------Step 2:
> x<-c(1,0)
> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

   B <  R [ 2 ] < A

 0.25 < 1.142857 < 4   -->  Countinue sampling.
> #-------------------------------------- --------------------------------------------------------Step 3:
> x<-c(1,0,0)
```

> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

   B < R [3] < A

 0.25 < 0.6530612 < 4   --> Countinue sampling.

> #--- ---Step 4:

> x<-c(1,0,0,1)

> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

   B < R [4] < A

 0.25 < 1.306122 < 4   --> Countinue sampling.

> #--- ---Step 5:

> x<-c(1,0,0,1,1)

> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

   B < R [5] < A

 0.25 < 2.612245 < 4   --> Countinue sampling.

 > #--Step 6:

> x<-c(1,0,0,1,1,1)

> sprt(dbinom(x,1,0.3),dbinom(x,1,0.6),0.2,0.2)

 R [6] > A

 5.22449 > 4   --> Rejcet H0.Finished

curve (N, R_N)
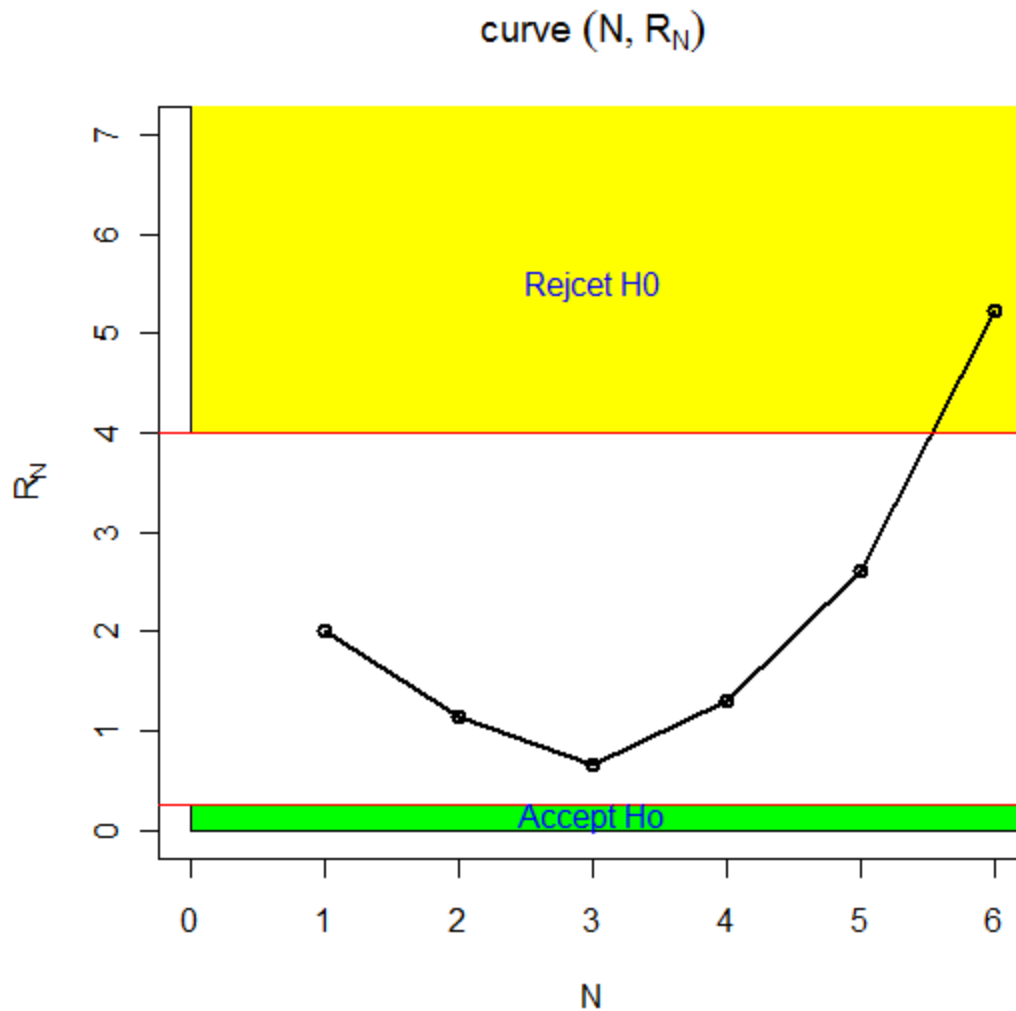
We repeatedly add the observed value to each step and substitute it in the density under H_0 and H_1. Then if we put the values of densities and the probability of statistical errors as the `sprt` function's inputs, sampling ends, like the example, at a particular step. In the final step, the path diagram (N, R_N) will be displayed automatically.

Problem 2

In many cases, the SPR test requires, on average, only half of the sample needed for the most powerful test (MPT). In other words, for each given pair (α, β), the sample size required in the sequential analysis is half of the sample size needed for the analysis based on fixed sample size; Therefore, we want to calculate the expected number of samples required under H_0 and H_1 in this method.

Suppose $X_1, X_2, ...$ is a sequence of independent random variables with the same probability density function (or probability function) f. We want to perform the following test using the SPRT method.

If $a = \ln(A) = \ln\left(\frac{1-\beta}{\alpha}\right)$, $b = \ln(B) = \ln\left(\frac{\beta}{1-\alpha}\right)$, and $z_1 = \ln\left(\frac{f_1(X_1)}{f_0(X_1)}\right)$ then:

$E_{H_0}(N) \approx \frac{a\alpha + b(1-\alpha)}{E_{H_0}(z_1)}$, and $E_{H_1}(N) \approx \frac{b\beta + a(1-\beta)}{E_{H_1}(z_1)}$

If $E_{H_0}(z_1) = 0$ or $E_{H_1}(z_1) = 0$ then:

$$E_{H_i}(N) \approx -\frac{ab}{E_{H_i}(z_1^2)}, \qquad i = 0,1$$

Example 2

Let $X_1, X_2, ...$ is a sequence of random variables of normal distribution with mean parameter θ and variance parameter $\sigma^2 = 100$. If $\alpha = 0.01$, $\beta = 0.05$ consider the problem of SPR testing

$$H_0: \theta = 100 \quad vs \quad H_1: \theta = 105$$

Programming 2

> # B .Falahi Fard 881537

> # The estimate of sample number for H0 & H1 in the SPRT.

> # Introduction to Theory of Statistics By: M .Mood, A .Graybill, C .Boes –

># Translated by :A.Meshkani , Issue:10 , Page:558

> # dname0 : We put the letter "d" befor acronym of the distribiution letter unther H0.

> # p0 : The parameters vector unther the H0 distribiution.

> # x0   : The vector of simulated data under the distribution of H0.

> # dname1 : We put the letter "d" befor acronym of the distribiution letter unther H1.

> # p1 : The parameters vector unther the H1 distribiution.

> # x1   : The vector of simulated data under the distribution of H1.

> # alpha: Probability type I error.

> # beta : Probability type II error . #--

> en.sprt<- function (dname0,p0,x0,dname1,p1,x1,alpha,beta){

```
+ a<-  log((1-beta)/alpha)

+ b<-  log(beta/(1-alpha))

+ if(length(p0)==3){

+ zH0<- log(dname1(x0,p1[1],p1[2],p1[3])/dname0(x0,p0[1],p0[2],p0[3]));

+ zH1<- log(dname1(x1,p1[1],p1[2],p1[3])/dname0(x1,p0[1],p0[2],p0[3]))}

+ else if(length(p0)==2){

+ zH0<- log(dname1(x0,p1[1],p1[2])/dname0(x0,p0[1],p0[2]));

+ zH1<- log(dname1(x1,p1[1],p1[2])/dname0(x1,p0[1],p0[2]))}

+ else if(length(p0)==1){

+ zH0<- log(dname1(x0,p1[1])/dname0(x0,p0[1]));

+ zH1<- log(dname1(x1,p1[1])/dname0(x1,p0[1]))}

+ if(mean(zH0)==0)EH0N<- -(a*b)/mean((zH0)^2)

+ else EH0N<- (a*alpha +b*(1-alpha))/mean(zH0)

+ if(mean(zH1)==0)EH1N<- -(a*b)/mean((zH1)^2)

+ else EH1N<- (a*(1-beta) + b*beta)/mean(zH1)

+ cat("\n","The estimate of  sample number for H0 :",EH0N,"\n",

+ "The estimate of  sample number for H1 :",EH1N,"\n","\n")

+ }#---------------------------------------------- Solve --------------------------

> x0<- rnorm(100000,100,10)

> x1<- rnorm(100000,105,10)

> en.sprt(dnorm,c(100,10),x0,dnorm,c(105,10),x1,0.01,0.05)

 The estimate of  sample number for H0 : 22.92658

 The estimate of  sample number for H1 : 33.3086
```

References

[1] Wald. A (1945). Sequential Tests of Statistical Hypotheses. Annals of Mathematical Statistics. 16 (2): 117–186.

[2] Mood. A, Graybill. F, Boes.D. (1973) Introduction to the Theory of Statistics. McGraw-Hill