Savage-Dickey

Manuel Villarreal-Ulloa

24/7/2019

Simulation

The objective of this simulation excersice is to test the behavior of the Savage-Dickey posterior density ratio test in a simple setting. First we will start by drawing samples from two Binomial distributions with the same probability parameter, then we will compare the bayes factors using a Beta conjugate prior and the SD method.

```
# Both samples in are drawn from the same distribution with an n value of 50
# and a probability of 0.5.
x1<-rbinom(1,50,0.5)
x2<-rbinom(1,50,0.5)
```

Posterior probability of equal probability model

Given that the Beta distribution is a conjugate prior for the Binomial, the posterior distribution of the model that assumes that samples are generated from a single model (one parameter), the posterior follows a Beta distribution with new parameters $\alpha^* = \alpha_p + \sum_i x_i$ and $\beta^* = \beta_p + n - \sum_i x_i$ which is also known as a Beta-Binomial distribution. We will compute the Bayes factor analitically using the ratio between integrated posterior likelihoods.

```
# Values of alpha and beta for the prior distribution.
ap<-1
bp<-1
# Integrated posterior likelihood for the one ratio model.
h0<-beta(x1+x2+ap,50-x1+50-x2+bp)
```

Posterior probability of different probability model

For the un-equal rates model, we eill assume that the samples from the two populations are drawn independently from two binomial distributions with parameters θ_1 and θ_2 and sample size $n_1 = n_2 = 50$. In this case, the integrated likelihood is the product of the Beta functions with the corresponding posterior α and β parameters. The Bayes Factor then is theratio of the integreted likelihood for the one distribution model and the one distribution model,which in this case should be lower than 1.

```
# Prior values for alpha and beta corresponding to each of the distributions.
ap_1<-1
bp_1<-1
ap_2<-1
bp_2<-1
# Integrated posterior likelihood for the two distributions model.
h1<-beta(x1+ap_1,50-x1+bp_1)*beta(x2+ap_2,50-x2+bp_2)
# Computing the analityc Bayes factor
BF<-h1/h0
```

The Bayes Factor under the assumption that the samples come from either one distribution or two, jas a value of 0.34 or 2.983 in favor of the one distribution model.

Calculating the Savage-Dickey aproximation

For the Savage-Dickey (SD) approximation, we need to calculate the ratio of the posterior distribution to the prior distribution of a transformation of the probability at the value of 0. In the case, of a Beta conjugate prior, the prior and posterior of the diference between probabilities might sometimes not have a closed form, however, Pham-Gia, Turkkan and Eng (1993) found a closed form for this transformation under some restrictions in the prior and posterior values of α and β. Under this setting, the density at 0 of the difference between probabilities is just

$$f(0) = \frac{B(\alpha_1 + \alpha_2 - 1, \beta_1 - \beta_2 - 1)}{B(\alpha_1, \beta_1)B(\alpha_2, \beta_2)} \tag{1}$$

In order to compute the SD approximation we only need the ratio of 1 using the posterior and prior values of α and β.

```
# Prior and Posterior density at 0
h0<-beta(ap_1+ap_2-1,bp_1+bp_2-1)/
    (beta(ap_1,bp_1)*beta(ap_2,bp_2))
h1<-beta(ap_1+x1+ap_2+x2-1,bp_1+50-x1+bp_2+50-x2-1)/
  (beta(ap_1+x1,bp_1+50-x1)*beta(ap_2+x2,bp_2+50-x2))
# Computing the Savage-Dickey
sd<-h1/h0
```

Using the SD approximaion and the density in equation 1 the bayes factor assosiated with the equal probabilities model is exactly the same with a factor of 2.983 in favor of the equality between θ_1 and θ_2.

This first example starts with the simulation of two variables with the same probability distribution and wiht both procedures we end up with the same conclussion, that the probability of a success in each population is the same. For the next example, we will use both methods but simulating the data using two distributions.

```
# Samples are drawn from two different distributions, one with a succes
# probability of 0.45 and the other with a probability of 0.55. we will
# keep the sample size equal between samples.
x1<-rbinom(1,50,0.4)
x2<-rbinom(1,50,0.6)
# One prob model
ap<-1
bp<-1
h0<-beta(x1+x2+ap,50-x1+50-x2+bp)
# Two probabilites model
ap_1<-1
bp_1<-1
ap_2<-1
bp_2<-1
h1<-beta(x1+ap_1,50-x1+bp_1)*beta(x2+ap_2,50-x2+bp_2)
# Bayes factor
BF<-h1/h0
# SD approximation
h0<-beta(ap_1+ap_2-1,bp_1+bp_2-1)/
    (beta(ap_1,bp_1)*beta(ap_2,bp_2))
h1<-beta(ap_1+x1+ap_2+x2-1,bp_1+50-x1+bp_2+50-x2-1)/
```

```
  (beta(ap_1+x1,bp_1+50-x1)*beta(ap_2+x2,bp_2+50-x2))
# Savage-Dickey
sd<-h1/h0
```

Again, both methods return the same result, however, in this case the conclusion is that there are two distributions, in other words, that θ_1 is different from θ_2 with a value of 1.876 and 1.876 respectively.