Computational Statistics-Report

David Niederkofler, Erlend Lokna

2022-12-01

```
mydata<-read.table("Report2_Dataset.txt", header=FALSE)
```

Statistical Analysis of Covariates

It is important to mention the use of notation before we proceed. We will in this section use the notation $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ as our covariate vector.

Bernoulli - Beta, Ascicles

Since the Ascicles - covariate has a 0-1 outcome we can assume that it is Bernoulli distributed with parameter θ. A natural conjugate prior for the Bernoulli distribution is the Beta distribution. Therefore we have the following initial information:

$$f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

$$L(\mathbf{x}|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$h(\theta) = Beta(a, b)$$

We proceed by calculating the posterior distribution for θ

$$h(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)h(\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}I(0<\theta<1)$$

$$\propto Beta(\theta|a + \sum_{i=1}^{n} x_i, b + n - \sum_{i+1}^{n} x_i)$$

1.1 Model selection

1.2 Results

1. Sex

The sex of the patients is encoded in a binary variable, where 0 means *male* and 1 means *female*.

1.1 Model selection

We assume a Bernoulli model $Ber(\theta)$ for the sex of the patient conditional on one parameter θ, the probability of the patient to be female. The density function is given by

$$f(x|\theta) = \theta^{x}(1-\theta)^{1-x}, \tag{1}$$

where $x \in \{0, 1\}$. As a prior distribution for θ we use the natural conjugate family of the Bernoulli distribution, namely the Beta distribution, $Beta(a, b)$, with two shape parameters $a = b = 2$ to give more weight to the middle of the interval $[0, 1]$, knowing how females and males are represented in the general population.

1.2 Results

From the given dataset we get the sample size n and the sum of the observations s:

```
n<-length(mydata$V6[!is.na(mydata$V6)])
s<-sum(mydata$V6)
n
```

```
## [1] 312
```

```
s
```

```
## [1] 276
```

Therefore the posterior distribution is $Beta(2+s, 2+n-s)$, which turns out to be $Beta(278, 38)$. From that we get

```
## Posterior mean:  0.8797468
```

```
## Posterior mode:  0.8821656
```

```
## Posterior 95% Confidence Interval: [ 0.8417454 , 0.9132003 ]
```

And the HPD confidence Interval calculates to:

```
tst<-rbeta(1e5,278,38)
hdi(tst)
```

```
##     lower     upper
## 0.8433167 0.9141901
## attr(,"credMass")
## [1] 0.95
```