BDA - Assignment 2

Anonymous

7/3/2020

Q - Inference for binomial proportion

We use the dataset algae.txtw which contains observations form 274 Finnish Lakes, where a "0" means: NO algae in the lake, and "I" means: Algae in lake.

a)

The unknown value π according to the observations (n=observations =274, y=lakes with algaes = 44) and prior knowledge:

$$p(\pi \mid y) \propto Beta(\alpha + y,\, \beta - y + n) \;=\; Beta(2 + 44,\, 10 + 274 - 44) \;=\; Beta(46,\, 240)$$

We now want to create a model "beta_point_est" that can calculate the posterior mean

```
# point estimate

beta_point_est <- function(prior_alpha, prior_beta, data){
  n <- length(data) # 274 observation
  y <- length(which(data==1)) # 44 lakes with algaes
  post_mean=(prior_alpha+y)/(prior_alpha+prior_beta+n)
  return(post_mean)
}

beta_point_est(prior_alpha = 2, prior_beta = 10, data = algae)
```

```
## [1] 0.1608392
```

We find the posterior mean to be 16.08 %

We now want to create a model "beta_interval" that can find the 90 % posterior interval (credible interval)

```
# posterior interval using qbeta (quantile function)

beta_interval <- function(prior_alpha, prior_beta, data, prob){
  n <- length(data)
  y <- length(which(data==1))
  pi_low <- qbeta((1-prob)/2, prior_alpha+y, prior_beta+n-y)
  pi_high <- qbeta(prob+(1-prob)/2, prior_alpha+y, prior_beta+n-y)
  post_int <- list(pi_low, pi_high)
  return(post_int)
```

```
}

beta_interval(prior_alpha = 2, prior_beta = 10, data = algae, prob = 0.9)
```

```
## [[1]]
## [1] 0.1265607
##
## [[2]]
## [1] 0.1978177
```

90% posterior interval: 12.66% - 19.78%

b)

Finding the probability that the proportion of monitoring sites with detectable algae levels π is smaller than $\pi(0) = 0.2$ that is known from historical records by creating function "beta_low"

```
# creating function beta_low with pbeta (distribution function):

beta_low <- function(prior_alpha, prior_beta, data, pi_0){
  n <- length(data)
  y <- length(which(data==1))
  dis_fun <- pbeta(pi_0, prior_alpha+y, prior_beta+n-y)
  return(dis_fun)
}

beta_low(prior_alpha = 2, prior_beta = 10, data = algae, pi_0 = 0.2)
```

```
## [1] 0.9586136
```

We find that the probability of that the algae level is smaller than the historical records is 95.86%

c)

Assumptions for the model used in b):

- Since we use β functions for the priors, they must follow a β distribution

Add more…

d)

Prior sensitivity analysis:
By testing different reasonable priors a sensitivity analysis will be carried out

```
# uniform prior (1, 1):
post_mean_1.1 <- beta_point_est(prior_alpha = 1, prior_beta = 1, data = algae) # posterior mean
post_int_1.1 <- beta_interval(prior_alpha = 1, prior_beta = 1, data = algae, prob = 0.9) # 90% posterio
post_mean_1.1; post_int_1.1
```

```
## [1] 0.1630435
```

```
## [[1]]
## [1] 0.1279681
##
## [[2]]
## [1] 0.2008987
```

```r
# uniform prior (2, 2) - weak/uninformative prior:
post_mean_2.2 <- beta_point_est(prior_alpha = 2, prior_beta = 2, data = algae) # # posterior mean
post_int_2.2 <- beta_interval(prior_alpha = 2, prior_beta = 2, data = algae, prob = 0.9) # 90% posterio
post_mean_2.2; post_int_2.2
```

```
## [1] 0.1654676
```

```
## [[1]]
## [1] 0.1302856
##
## [[2]]
## [1] 0.2033897
```


```r
# uniform prior (2, 2) - weak/uninformative prior:
post_mean_2.2 <- beta_point_est(prior_alpha = 2, prior_beta = 2, data = algae) # # posterior mean
post_int_2.2 <- beta_interval(prior_alpha = 2, prior_beta = 2, data = algae, prob = 0.9) # 90% posterio
post_mean_2.2; post_int_2.2
```

```
## [1] 0.1654676
```

```
## [[1]]
## [1] 0.1302856
##
## [[2]]
## [1] 0.2033897
```


```r
# String prior (50, 150) - weak/uninformative prior:
post_mean_50.150 <- beta_point_est(prior_alpha = 50, prior_beta = 150, data = algae) # # posterior mean
post_int_50.150 <- beta_interval(prior_alpha = 50, prior_beta = 150, data = algae, prob = 0.9) # 90% po
post_mean_50.150; post_int_50.150
```

```
## [1] 0.1983122
```

```
## [[1]]
## [1] 0.1689568
##
## [[2]]
## [1] 0.2291163
```

With a weak prior the data is dominant in the posteior.
With a strong prior the posterior interval is more narrow.