Bayesian Hypothesis Testing

BIOS719 Generalized Linear Models

Testing hypothesis

An investigator arrives at a village and after observing the villagers' contact with water, she is 80% sure that infection will be endemic in this village. The investigator takes stool samples from 10 villagers and 7 are positive for the parasite.

H0: Infection is not endemic ($\theta \leq 0.5$)

H1: Infection is endemic ($\theta > 0.5$)

Table 12.1 *Calculations for updating the investigator's priors for H_0 and H_1 to posteriors using the observed data on positive samples for* Schistosoma japonicum.

θ	Hypothesis	$P(\theta)$ Prior	$P(y\|\theta)$ Likelihood	$P(y\|\theta) \times P(\theta)$ Likelihood×Prior	$P(\theta\|y)$ Posterior
0.0	H_0	0.0333	0.0000	0.0000	0.0000
0.1	H_0	0.0333	0.0000	0.0000	0.0000
0.2	H_0	0.0333	0.0008	0.0000	0.0002
0.3	H_0	0.0333	0.0090	0.0003	0.0024
0.4	H_0	0.0333	0.0425	0.0014	0.0114
0.5	H_0	0.0333	0.1172	0.0039	0.0315
Sum		0.2000			0.0455
0.6	H_1	0.1600	0.2150	0.0344	0.2771
0.7	H_1	0.1600	0.2668	0.0427	0.3439
0.8	H_1	0.1600	0.2013	0.0322	0.2595
0.9	H_1	0.1600	0.0574	0.0092	0.0740
1.0	H_1	0.1600	0.0000	0.0000	0.0000
Sum		0.8000		0.1241	0.9545

```
f.table12.1 <- function(theta.cut, prior.H1, bysize, tab=T) {
  theta <- seq(0, 1, by=bysize)
  H1.theta <- ifelse(theta > theta.cut,1,0)   ## H1

  theta.prior <- ifelse(H1.theta==1,
                        prior.H1/length(H1.theta[H1.theta==1]),
                        (1-prior.H1)/length(H1.theta[H1.theta==0]))
  lik <- dbinom(x=7, size=10, prob=theta)
  lik.theta.prior <- lik*theta.prior
  theta.post <- lik.theta.prior/sum(lik.theta.prior)

  out <- cbind(theta=theta, H1=H1.theta, theta.prior=theta.prior, lik=lik,
               lik.theta.prior=lik.theta.prior, theta.post=theta.post)

  if(tab==T) {
    return(list(tab12.1=round(out, 4),
                H.prior=tapply(theta.prior, H1.theta, sum),
                H.post=tapply(theta.post, H1.theta, sum)))
  } else {
    return(list(H.prior=tapply(theta.prior, H1.theta, sum),
                H.post=tapply(theta.post, H1.theta, sum),
                Equal.tail.interval=
                  c( max(theta[cumsum(theta.post)<=0.05/2]), min(theta[cumsum(theta.post)>1-0.05/2])),
```

```
                theta=theta,
                theta.post=theta.post))
  }
}

test1 <- f.table12.1(theta.cut=0.5, prior.H1=0.8, bysize=0.1, tab=T)
test1
```

```
## $tab12.1
##       theta H1 theta.prior    lik lik.theta.prior theta.post
##  [1,]   0.0  0      0.0333 0.0000          0.0000     0.0000
##  [2,]   0.1  0      0.0333 0.0000          0.0000     0.0000
##  [3,]   0.2  0      0.0333 0.0008          0.0000     0.0002
##  [4,]   0.3  0      0.0333 0.0090          0.0003     0.0024
##  [5,]   0.4  0      0.0333 0.0425          0.0014     0.0114
##  [6,]   0.5  0      0.0333 0.1172          0.0039     0.0315
##  [7,]   0.6  1      0.1600 0.2150          0.0344     0.2771
##  [8,]   0.7  1      0.1600 0.2668          0.0427     0.3439
##  [9,]   0.8  1      0.1600 0.2013          0.0322     0.2595
## [10,]   0.9  1      0.1600 0.0574          0.0092     0.0740
## [11,]   1.0  1      0.1600 0.0000          0.0000     0.0000
##
## $H.prior
##   0   1
## 0.2 0.8
##
## $H.post
##          0          1
## 0.04550201 0.95449799
```

```
test1$H.prior
```

```
##   0   1
## 0.2 0.8
```

```
test1$H.post
```

```
##          0          1
## 0.04550201 0.95449799
```

```
BF1 <- (test1$H.post[2]/test1$H.post[1])/(test1$H.prior[2]/test1$H.prior[1])
BF1
```

```
##        1
## 5.244263
```

```
test2 <- f.table12.1(theta.cut=0.5, prior.H1=0.5, bysize=0.0001, tab=F)
test2$H.prior
```

```
##   0   1
## 0.5 0.5
```

```
test2$H.post
```

```
##         0         1
## 0.1133256 0.8866744
```

```r
BF2 <- (test2$H.post[2]/test2$H.post[1])/(test2$H.prior[2]/test2$H.prior[1])
BF2
```

```
##       1
## 7.82413
```

```r
#### Frequentist test
prop.test(x=7, n=10, p=0.5, alternative=c("greater"), correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  7 out of 10, null probability 0.5
## X-squared = 1.6, df = 1, p-value = 0.103
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.4416998 1.0000000
## sample estimates:
##   p
## 0.7
```