**5380 Homework 2**

15 February 2021

Professor Wulff

Cullen Blair

* Show work on derivation problems and give R code for computation problems.
* Posterior summaries are posterior mean, standard deviation, 95% equal tails interval.
* Use set.seed(0) when taking 100000 samples from the respective distributions.

P.2.1 Consider P.3.10.4 (p 80). Let *j=0* index the control and *j=*1 index the treatment. Consider the population odds ratio where P*j* iid Beta (1,1)

1. Identify the exact posterior distributions for Po, P1 (use results on p 35). Take samples from the posterior distributions to obtain Bayesian summaries for Po, P1,  . Interpret the Bayes estimate for this application.

Beta(

samp.out=function(x) c(mean=mean(x),sd=sd(x),quantile(x,probs=c(0.025,0.975)))  
n.ctrl=674;n.trt=680; y.ctrl=39;y.trt=22;a=1;b=1  
set.seed(0)  
n.sim=100000  
sample.1a.ctrl=rbeta(n.sim,y.ctrl+a,n.ctrl-y.ctrl+b)  
sample.1a.trt=rbeta(n.sim,y.trt+a,n.trt-y.trt+b)  
theta.h=(sample.1a.trt/(1-sample.1a.trt))\*((1-sample.1a.ctrl)/sample.1a.ctrl)  
vec=cbind(sample.1a.ctrl,sample.1a.trt,theta.h)  
t(apply(vec,2,samp.out))

## mean sd 2.5% 97.5%  
## sample.1a.ctrl 0.05914660 0.009045640 0.0425253 0.07795534  
## sample.1a.trt 0.03370162 0.006882446 0.0215396 0.04846575  
## theta.h 0.56992413 0.155338250 0.3217225 0.92736398

The Bayes estimate of the population mean probability of death for the control group (Po) is 0.05914660 which corresponds to the posterior mean. The Bayes estimate of the population mean probability of death for the treatment group (P1) is 0.03370162 which corresponds to the posterior mean. The Bayes estimate of the population mean odds ratio of death for those who receive the beta blocker versus those who do not () is 0.56992413 which corresponds to the posterior mean.

1. Identify the approximate posterior distributions for Po, P1 (use results from class on 02/09). Take samples from the posterior distributions to obtain Bayesian summaries for Po, P1,  . Interpret the 95% Bayes credible interval for this application.

By Bayesian CLT, )

set.seed(0)  
samp.out.var=function(x) c(mean=mean(x),sd=sd(x),quantile(x,probs=c(0.025,0.975)))  
ybar.ctrl=y.ctrl/n.ctrl  
ybar.trt=y.trt/n.trt  
sample.1b.ctrl=rnorm(n.sim,ybar.ctrl,sqrt((ybar.ctrl\*(1-ybar.ctrl))/n.ctrl))  
sample.1b.trt=rnorm(n.sim,ybar.trt,sqrt((ybar.trt\*(1-ybar.trt))/n.trt))  
theta.h.b=(sample.1b.trt/(1-sample.1b.trt))\*((1-sample.1b.ctrl)/sample.1b.ctrl)  
vec2=cbind(sample.1b.ctrl,sample.1b.trt,theta.h.b)  
t(apply(vec2,2,samp.out.var))

## mean sd 2.5% 97.5%  
## sample.1b.ctrl 0.05787618 0.009002765 0.04020622 0.07548760  
## sample.1b.trt 0.03232407 0.006765440 0.01901993 0.04560662  
## theta.h.b 0.55964631 0.158411462 0.29618965 0.91800212

There is 95% posterior probability that the population mean probability of death of control group patients (Po) is in [0.04020622,0.07548760] which corresponds to the equal tails interval. There is 95% posterior probability that the population mean probability of death of treatment group patients (P1) is in [0.01901993,0.04560662] which corresponds to the equal tails interval. There is 95% posterior probability that the population mean odds ratio of death for those that receive the beta blocker treatment versus those who don’t () is in [0.29618965,0.91800212] which corresponds to the equal tails interval.

1. Compute the Bayesian p-value for testing H0: , H1: using either (a) or (b). Explain whether the treatment leads to a decrease in the population odds of death according to the approximate Jeffrey measure of evidence.

length(theta.h[theta.h>=1])/length(theta.h)

## [1] 0.01238

Using the output from (a) the p-value of 0.01238 corresponds to the probability that , resulting in substantial evidence against the claim that the treatment leads to an increase in the population odds of death.

P.2.2 Suppose a bowl contains red, blue, and white chips where *m* chips 7 are selected at random with replacement. The true proportions of red and blue chips are unknown. Let *Xj* denote the number of chips from color category *j* out of *m*. Assume ~ Multinomial () and ~Dirichlet(*a,b,c*).

1. Derive and identify the joint posterior distribution . Give the updated hyperparameter values *a\*, b\*, c\** for this conjugate family.

*~ Dirichlet ( , ) ~ Dirichlet (a\*, b\*, c\*)*

1. Derive and identify the marginal posterior distribution .. Integrate with correct limits, integral substitution , and Beta distribution properties.

~Beta()

1. Obtain the prior and posterior summaries of and when a=2, *b*=2, c=4, *m*=6, *x1*=3, *x2*=0. Carefully explain the differences between this informative prior distribution and the posterior distribution based upon each of these summaries.

library(DirichletReg)

## Loading required package: Formula

set.seed(0)  
a=2;b=2;c=4;m=6;x1=3;x2=0  
new.a=a+x1;new.b=b+x2;new.c=c+m-x1-x2  
prior.theta=rdirichlet(n.sim,c(a,b,c))  
colnames(prior.theta)=c("theta1","theta2","theta3")  
t(apply(prior.theta,2,samp.out))

## mean sd 2.5% 97.5%  
## theta1 0.2504995 0.1449129 0.03634544 0.5791604  
## theta2 0.2490246 0.1439369 0.03688625 0.5765323

## theta3+theta1+theta0 sum to 1 as a check.

post.theta1=rbeta(n.sim,new.a,new.b+new.c)  
post.theta2=rbeta(n.sim,new.b,new.a+new.c)  
post.thetas=cbind(post.theta1,post.theta2)  
t(apply(post.thetas,2,samp.out))

## mean sd 2.5% 97.5%  
## post.theta1 0.3571167 0.12372304 0.13811779 0.6135281  
## post.theta2 0.1427649 0.09026525 0.01907784 0.3587090

🡺 ~Beta()

Re-integrating the joint posterior I found the marginal posterior distribution for .

When comparing the summaries of the prior and posterior summaries of and , it is worth noting how similar the prior summaries are between and . This is due to the fact that their summaries are from the same distribution. However, in the posterior summaries of and the estimates and uncertainties are not nearly as similar. I believe that the posterior summaries give a more fine tuned picture of each parameter as both the prior and the sampling distributions are considered in their respective posterior distributions.