Bayesian Poisson regression model and Breast cancer and the age at which she has the first child

Honya Elfayoumy

ISYE 6420 Final Project

1. Introduction

There are many studies that have obtained a relationship between breast cancer and the age of a woman's first pregnancy. For example we can see that in the following study https://www.cell.com/cell-stem-cell/fulltext/S1934-5909(13)00197-5, the researchers looked at the activity of specific cells in breast tissue samples from thousands of women that were obtained over the course of 20 years.

The researchers found that women who had a full-term pregnancy in their 20s and 30s had fewer numbers of certain breast cells called "mammary gland progenitor cells." These cells can divide and become milk-producing cells. Full-term pregnancies before age 30 also reduced the ability of progenitor cells in the mammary glands to grow and divide. If the cells cannot grow and divide, they are less likely to mutate and lead to breast cancer.

By comparing many samples of breast cancer tissue, the researchers concluded that women at high risk for breast cancer, such as women with a BRCA1 or BRCA2 gene abnormality, have higher amounts of stem cells from the mammary glands than the number average.

Overall, women who had a full-term pregnancy in their 20s and 30s had the fewest stem cells from the mammary glands, even compared to women who had never been diagnosed with breast cancer and who had never been pregnant. . However, women who gave birth before age 30 and who later developed breast cancer had a greater than average number of stem cells from the mammary glands.

2. Data Set

We are considering data of breast cancer and the age at which she has the first child (see Mahon et al. (1970) Bulletin of the World Health Organization). We can find here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2427645/pdf/bullwho00206-0003.pdf.

In this data we can see if the woman has breast cancer and the age at which she has the first child in several hospitals of USA, Greece, Yugoslavia, Brazil, Taiwan and Japan.

Therefore we can summarize the data in the following 2×2 contingency table:

	Age at first birth	
Status	Age $\geq$ 30 (1)	Age < 30 (0)

Cancer (1)	683	2537
Cancer (0)	1498	8747

Therefore performing a manual count of the data obtained in the paper, we managed to obtain the following results:

Status	**Age**	**Counts**
1	1	683
1	0	2537
0	1	1498
0	0	8747

Where

- Status=1 the woman got cancer, Status=0 the woman not got cancer
- Ages=1 the age of the woman is over 30 years old in her first pregnancy, Ages=0 the age of the woman is less than 30 years old in her first pregnancy
- Count: represents the number of cases in the 4 possible combinations

3. Objective

The main objective of this project is to find a Bayesian model that allows us to verify or refute what medical research assures, "The risks of developing breast cancer are greater in women whose first pregnancy occurs after the age of 30". Because the nature of the data are counts, we will use a Bayesian Poisson model, and we will study the interaction between the variable Status and age, and thus we will see that if this variable is greater than one then the presence of the factor (age) is associated the greater the occurrence of the event (breast cancer).

The model

We want to fit a Bayesian Poisson regression model for the number of breast cancer cases for women who had their first children after age 30 or before

$\text{count}_i \sim Poisson(\lambda_i)$

$log(\lambda_i) = \mu + a \times \text{status}_i + b \times \text{ages}_i + \boldsymbol{ab} \times \mathbf{status}_{\boldsymbol{i}} \times \mathbf{ages}_{\boldsymbol{i}}$

Suppose the following prior distributions are placed on the parameters

- $\pi(a) \sim \text{Normal}(0, \sigma^2 = 1)$
- $\pi(b) \sim \text{Normal}(0, \sigma^2 = 1)$
- $\pi(ab) \sim \text{Normal}(0, \sigma^2 = 1)$

Running the model we obtain the following result:

	mean	sd	2.5%	97.5%	Rhat	n.eff
a	-1.24	0.0224	-1.28	-1.19	1	2200
b	-1.76	0.0277	-1.81	-1.71	1	8000
ratio.odds	1.56	0.0810	1.41	1.73	1	3500
deviance	42.15	2.8193	38.58	49.18	1	5600

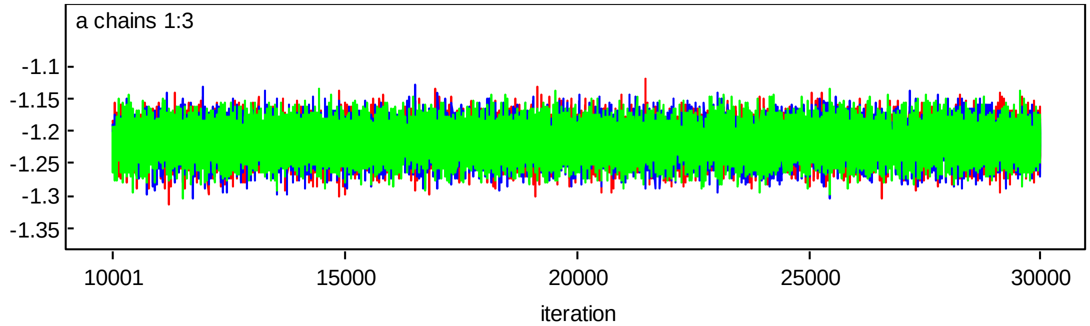
We obtain a very good model where

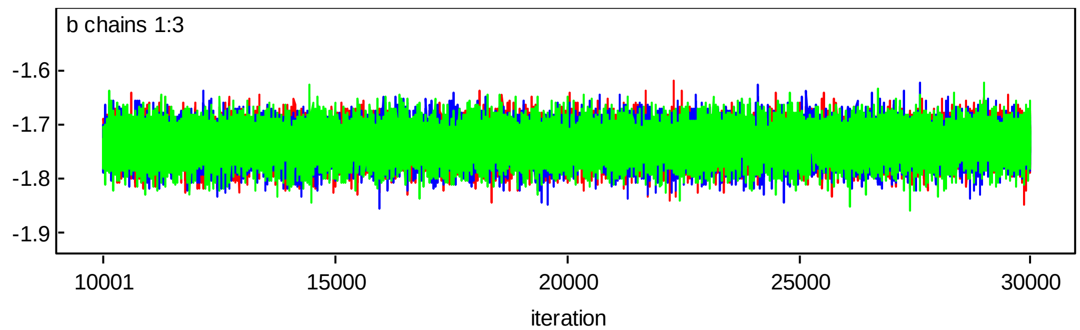
- $a = -1.24,$
- $b = -1.76,$
- $ab = \log(1.56)$

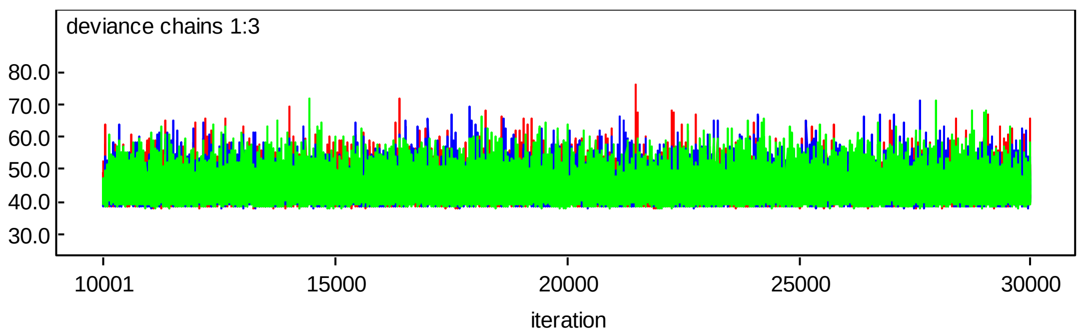
The 95% Credible interval of the slopes are:

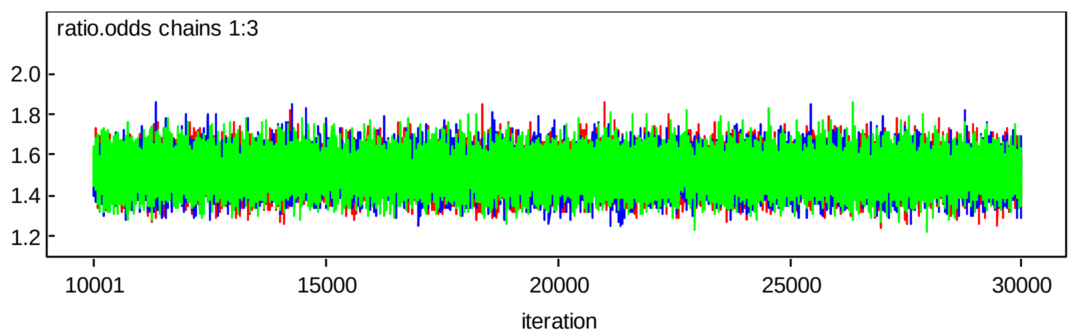
- For a we obtaing [-1.28; -1.19]
- For b we obtaing [-1.81; -1.71]
- For a we obtaing [1.41 ; 1.73]

We can note that credibility intervals does not contain zero so we can say that all variables are significant.









In the following figures we observe the trance plot, which reflects a good convergence of the method, the odd ratio of the variable **ab** (interaction between the variable Status and age) equal to 1.56 what is greater than one, which can be interpreted as a percentage of about 56% higher in cases of breast cancer, for people over 30 years.

Conclusion

A Bayesian Poisson regression model for the number of the number of breast cancer cases for women who had their first children after age 30 or before is performed and a relationship was found between age at first birth and risk of breast cancer that was to be expected from medical research. It is estimated that women who have their first child when they have more than 30 years have a 56% increased risk of breast cancer than those whose first birth is before age 30 years.

Appendix

The Model

```
write("model pois;
 {
 for(i in 1:N) {
   counts[i] ~ dpois(lambda[i])
   log(lambda[i]) <- mu+a*status[i] + b*ages[i]+ab*status[i]*ages[i]
 }
   ratio.odds <-exp(ab)
   mu ~ dnorm (0.0, 1)
   a ~ dnorm (0.0, 1)
   b ~ dnorm (0.0, 1)
  ab ~ dnorm (0.0, 1)
 }
 ","modelPois.txt")
```

Runing the Model:

```
status = c(1,1,0,0)
ages = c (1,0,1,0)
counts = c (683,2537,1498,8747)
N=length(counts)

data=list("N","status","ages","counts")
params=c("a","b","ratio.odds" )
inits <- function () { list ( mu = rnorm (1) ,a = rnorm (1),b = rnorm (1)
,ab = rnorm (1) ) }

nc <- 2      #number of MCMC chains to run
ni <- 5000  #number of samples for each chain
nb <- 1000   #number of samples to discard as burn-in
nt <- 1      #thinning rate, increase this to reduce autocorrelation

bugs.out <- bugs(data=data, inits=inits, parameters.to.save=params, model
.file="modelPois.txt",n.chains=nc, n.iter=ni, n.burnin=nb, n.thin=nt, deb
```

```
ug=TRUE, DIC=TRUE, bugs.directory = "C:\\Program Files\\WinBUGS14", worki
ng.directory=getwd())
```

References:

Study Suggests Why Giving Birth in 20s Reduces Breast Cancer Risk
docs/IntroWinBUGSwithR.md at master · BES-QSIG/docs (github.com)