# 732A73: Bayesian Learning

Computer Lab 2

*Oriol Garrobé, Dávid Hrabovszki*

*25 April 2020*

## Question 2. Posterior approximation for classification with logistic regression

The dataset WomenWork.dat contains n = 200 observations (i.e. women) on the following nine variables:

Constant, HusbandInc, EducYears, ExpYears, ExpYears2, Age, NSmallChild, NBigChild

### a

We consider the logistic regression

$Pr(y=1|x) = \frac{exp(x^T\beta)}{1+exp(x^T\beta)}$

where $y = 0$ is the woman does not work and $y = 1$ if she does. $X$ is an 8-dimensional vector with the features including the intercept. In this part, we approximate the posterior distribution of the parameter vector $\beta$:

$\beta|y, X \sim N\left(\tilde{\beta}, J_y^{-1}(\tilde{\beta})\right)$

$\tilde{\beta}$ is the posterior mode and $J(\tilde{\beta}) = -\frac{\partial^2 lnp(\beta|y)}{\partial\beta\partial\beta^T}|_{\beta=\tilde{\beta}}$ is the observed Hessian evaluated at the posterior mode. We use the following prior: $\beta \sim N(0, \tau^2 I)$, with $\tau = 10$.

The code for the approximation can be found and in the Appendix and was written using Mattias Villani's implementation as a template: https://github.com/mattiasvillani/BayesLearnCourse/raw/master/Code/MainOptimizeSpam.zip

We obtained the following posterior mode for $\beta$:

| Beta1 | Beta2 | Beta3 | Beta4 | Beta5 | Beta6 | Beta7 | Beta8 |
|---|---|---|---|---|---|---|---|
| 0.6267043 | -0.0197913 | 0.1802191 | 0.1675718 | -0.1446182 | -0.0820651 | -1.359151 | -0.024684 |

And the posterior covariance matrix $J_y^{-1}(\tilde{\beta})$:

| | Beta1 | Beta2 | Beta3 | Beta4 | Beta5 | Beta6 | Beta7 | Beta8 |
|---|---|---|---|---|---|---|---|---|
| Beta1 | 2.26603 | 0.00334 | -0.06545 | -0.01179 | 0.04578 | -0.03029 | -0.18875 | -0.09802 |
| Beta2 | 0.00334 | 0.00025 | -0.00056 | -0.00003 | 0.00014 | -0.00004 | 0.00051 | -0.00014 |
| Beta3 | -0.06545 | -0.00056 | 0.00622 | -0.00036 | 0.00190 | 0.00000 | -0.00613 | 0.00175 |
| Beta4 | -0.01179 | -0.00003 | -0.00036 | 0.00435 | -0.01425 | -0.00013 | -0.00147 | 0.00054 |
| Beta5 | 0.04578 | 0.00014 | 0.00190 | -0.01425 | 0.05558 | -0.00033 | 0.00321 | 0.00051 |
| Beta6 | -0.03029 | -0.00004 | 0.00000 | -0.00013 | -0.00033 | 0.00072 | 0.00518 | 0.00110 |
| Beta7 | -0.18875 | 0.00051 | -0.00613 | -0.00147 | 0.00321 | 0.00518 | 0.15126 | 0.00677 |
| Beta8 | -0.09802 | -0.00014 | 0.00175 | 0.00054 | 0.00051 | 0.00110 | 0.00677 | 0.01997 |

We also compute an approximate 95% credible interval for the coefficient of the NSmallChild variable: $[-2.121445, -0.5968567]$ NSmallChild is an important determinant of whether a woman is working or not, because it has the highest absolute valued posterior coefficient $\tilde{\beta}$ of all features. The posterior mode of its coefficient is -1.36, which means that the more small children a woman has, the less likely it is that she's working, because the sign is negative.

To verify that we got the correct result, we compare it to maximum likelihood estimates, and note that the numbers are very similar.

| Constant | HusbandInc | EducYears | ExpYears | ExpYears2 | Age | NSmallChild | NBigChild |
|---|---|---|---|---|---|---|---|
| 0.6443036 | -0.0197746 | 0.1798806 | 0.1675127 | -0.1443595 | -0.0823403 | -1.362502 | -0.0254299 |

**b**

In this part we simulate from the predictive distribution of the response variable in a logistic regression. More specificially we simulate and plot the predictive distribution for the Work variable for a 40 year old woman, with two children (3 and 9 years old), 8 years of education, 10 years of experience. and a husband with an income of 10. To do this, first we draw random numbers from the multivariate normal distribution, with mean $\tilde{\beta}$ and covariance matrix $J_y^{-1}(\tilde{\beta})$. Since in normal distributions the mean is equal to the mode, we can use the result from 2.a. Then we apply our implementation of the logistic regression function defined above to the given woman with every random draw.

The resulting distribution in Figue 1 shows that the given woman is much more likely to not be working, because the mode is around 0.2. This is mostly because she has a small child, which as we've seen before, is correlated negatively with working.

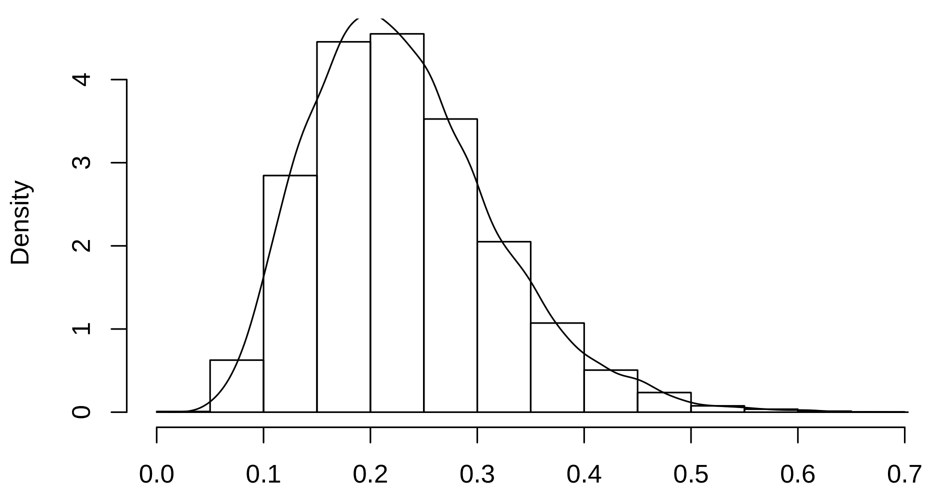


Figure 1: Histogram of the predictive distribution of the response variable of a woman with given features

**c**

Now, we consider 10 women which all have the same features as the woman in 2(b). The random variable of whether a woman is working or not follows a Bernoulli distribution, but if we add 10 such random variables, we get a Binomial distribution, with parameters n = 10 and p = p(given woman works). We calculate the probability of working as the mean of the Bernoulli random variables calculated in 2.b.

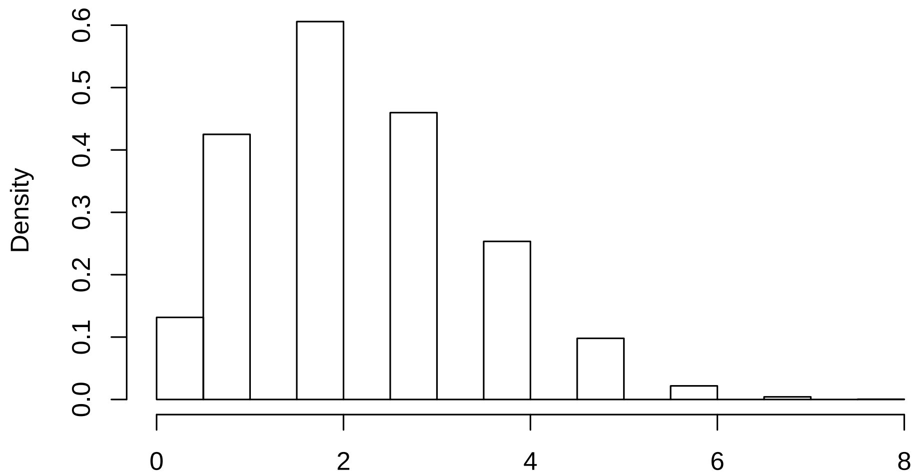


Figure 2: Histogram of the predictive distribution of the number of women working (out of 10)

Figue 2 shows that 2 such women working has the highest probability.

# Appendix

```
knitr::opts_chunk$set(echo=FALSE, eval=TRUE)
# R version
RNGversion('3.5.1')

# libraries
library(mvtnorm)
library(knitr)
# 2. Posterior approximation for classification with logistic regression

# 2.a

# The following code was written using Mattias Villani's implementation as a template:
# https://github.com/mattiasvillani/BayesLearnCourse/raw/master/Code/MainOptimizeSpam.zip

# read data
women = read.table('WomenWork.dat', header = T)

y = as.vector(women[,1])
x = as.matrix(women[,2:9])
tau = 10
nfeatures = dim(x)[2]

# prior
mu = as.vector(rep(0,nfeatures))
sigma = diag(tau^2, nfeatures, nfeatures)

LogPrior = function(Beta, mu, sigma){
  return(dmvnorm(Beta, as.matrix(mu), sigma, log = T))
}

# log likelihood
LogLikelihood = function(Beta, y, x){
  sum(y * x%*%Beta - log(1 + exp(x%*%Beta)))
}

# log posterior
LogPosterior = function(Beta, y, x, mu, sigma){
  LogPrior(Beta, mu, sigma) + LogLikelihood(Beta, y, x)
  # we can sum them instead of multiplying, because of the log
}

# initialize Beta vector randomly
# Seed
set.seed(1234567890)
Beta_init = as.vector(rnorm(dim(x)[2]))

# optimizing the log posterior by changing the Betas (maximize)
res = optim(Beta_init, LogPosterior, gr = NULL, y, x, mu, sigma,
                    method="BFGS", control=list(fnscale=-1), hessian=T)

Beta_hat = res$par # posterior mode
```

```r
Hessian = res$hessian
post_sigma = solve(-Hessian) # posterior cov matrix
stdev = sqrt(diag(post_sigma))

# approximate 95% credible interval for NSmallChild
lower = Beta_hat[7] - 1.96 * stdev[7] #-2.121445
upper = Beta_hat[7] + 1.96 * stdev[7] #-0.5968567

betas = t(as.matrix(Beta_hat))
colnames(betas) = c("Beta1","Beta2","Beta3","Beta4","Beta5","Beta6","Beta7","Beta8")
kable(betas)
sigma_table = data.frame(post_sigma)
colnames(sigma_table) = c("Beta1","Beta2","Beta3","Beta4","Beta5","Beta6","Beta7","Beta8")
rownames(sigma_table) = c("Beta1","Beta2","Beta3","Beta4","Beta5","Beta6","Beta7","Beta8")
kable(round(sigma_table,5))

# check results
glmModel <- glm(Work~0 + ., data = women, family = binomial)
kable(t(as.matrix(glmModel$coefficients))) #roughly the same as Beta_hat
# 2.b
# The mode of a normal distribution is equal to the mean.

log_reg = function(Beta, x){
  return(exp(x%*%Beta) / (1 + exp(x%*%Beta)))
}
Constant = 1
HusbandInc = 10
EducYears = 8
ExpYears = 10
ExpYears2 = (ExpYears/10)^2
Age = 40
NSmallChild = 1
NBigChild = 1

x_pred = c(Constant, HusbandInc, EducYears, ExpYears, ExpYears2, Age, NSmallChild, NBigChild)

# simulate posterior draws from Beta
n = 10000
post_sim = rmvnorm(n, mean = Beta_hat, sigma = post_sigma) # should i add pop variance to post_sigma?

# calculate target y
pred = apply(post_sim, 1, log_reg, x_pred)
hist(pred, freq = F, main = "", xlab = "")
lines(density(pred))
# 2.c
# probability of success (working)
p = mean(pred)

binom_sim = rbinom(10000, 10, p)
hist(binom_sim, freq = F, main = "", xlab = "")
```