Bayesian Data Analysis EC543

Instructor: M.A. Rahman

Submitted by: Mohit Shukla 11435

Date: April 3, 2015

Problem:

The data set attached refers to the serum concentration of immunoglobulin-G in 298 children aged from 6 months to 6 years.

Model 1: y1 = beta1 + beta2*x + e

Model2: y2 = beta1 + beta2*x + beta3*x^2 + e

(a) Plot a histogram and comment on skewness
(b) Estimate Model1 using ordinary least square. Comment of regression relationship.
(c) Estimate Model2 using ordinary least square. Comment on regression relationship.
(d) Bayesian Estimation: assume the errors to follow a standard normal distribution. Write down likelihood function and estimate the model using independent prior on beta and sigma square.

Solution:

Descriptive and exploratory analysis showed that age data is more skewed towards lower values and IgG is slightly skewed towards large values.

First model was a linear regression model, OLS estimation showed positive between IgG and age but R-squared was low which shows poor model fit of data.

Second model: OLS estimation showed a parabolic relation between age and IgG. R squared was larger than previous case that shows, this model was better.

Bayesian estimation of second model was done assuming errors to be normally distributed and independent prior for Beta and Sigma.

Code:

```
%Mohit Shukla
%%Q4.i

data=xlsread('C:/Users/caped_crusader/Bayesian Data
Analyses/Assignments/igg.xlsx');

age(1:298)=data(1:298,2);
igg(1:298)=data(1:298,3);

%Summary of age
mean_age=mean(age)
median_age=median(age)
std_age=std(age)
max_age=max(age)
min_age=min(age)

%Summary of Immunoglobulin-G
mean_igg=mean(igg)
median_igg=median(igg)
std_igg=std(igg)
max_igg=max(igg)
min_igg=min(igg)

%Histogram
hist(age,20);   %Histogram with 20 bins
%Inference: Data is skewed on both sides but the lower data points have
%high skewness

hist(igg,20);      %Histogram with 20 bins
%Inference: Data is slightly skewed towards large values


%%
%Q4. ii
N=298;
k=2;
x=ones(298,2);
x(1:298,2)=age(1:298);
y=ones(298,1);
y(1:298,1)=igg(1:298);
b_ols=(((transpose(x))*x)^-1)*(transpose(x)*y)   %OLS beta
sse=transpose(y-x*b_ols)*(y-x*b_ols);
mse=sse/(N-k)
sst=transpose(y-mean(y))*(y-mean(y));
r_sq=1-(sse/sst)

%Comment: From the regression we saw a positive relationship between IGg
%and age. But the R-squared is significantly low which implies that the
%model has not fit the data
```

```
%%
%Q4. iii
N=298;
k=3;
x=ones(298,3);
x(1:298,2)=age(1:298);
for i=1:298
x(i,3)=age(i)^2;
end
y=ones(298,1);
y(1:298,1)=igg(1:298);
b_ols=(((transpose(x))*x)^-1)*(transpose(x)*y)   %OLS beta
sse=transpose(y-x*b_ols)*(y-x*b_ols);
mse=sse/(N-k)
sst=transpose(y-mean(y))*(y-mean(y));
r_sq=1-(sse/sst)
%Comment: From the regression we saw a positive relationship between IGg
%and age but a negative relationship of IGg with age-squared. Which means
%that the relationship first increases and then decreases (parabolic)
%The R-squared is significantly low, though larger than previous case,
%which implies that the model has not fit the data but the fitting has
%improved after introducing x-square covariate.

%%
%Q4.iv
%Hyperparameters
n_post=N-n_pr;
y=data(1:298,1);   %Dependent Variable
x=ones(298,3);      %Covariate X1
x(1:298,2)=age(1:298);
for i=1:298
x(i,3)=age(i)^2;
end
b_ols=(((transpose(x))*x)^-1)*(transpose(x)*y);   %OLS beta
xsq=(transpose(x)*x);

V_pr=[1 0 0;0 2 0; 0 0 4];
b_pr=[0;1;1/6];       %Prior beta
h=1/2.77;
N=298;
n_pr=5;     %Prior nu; small weight to prior and more to likelihood
k=3;        %No. of parameters

var_b_pr=n_pr*V_pr/((n_pr-2)*h);        %Variance of prior beta
sd_b_pr=sqrtm(var_b_pr);                %Std Dev of prior beta

s_sq_inv=h^-1;
G=10000;
sum_b_exp=0;
for i=1:G
   V_posterior=(V_pr^-1 + h*xsq)^-1;
   b_posterior=V_posterior*(V_pr^-1*b_pr + h*transpose(x)*y); %Posterior
mean for ind. prior
   b=transpose(mvnrnd(b_posterior,V_posterior));        %Draw from dist
```

```
   n_posterior=N+n_pr;
   s_sq_inv=((transpose(y-x*b))*(y-x*b)+n_pr*(h^-1))/n_posterior; %Scale
for ind. h prior
   h=gamrnd(n_posterior/2,s_sq_inv);            %Draw of h from gamma
   if i>=0.1*G
   sum_b_exp=sum_b_exp+b;            %Acceptance after 1000 burns
   end
end

b_exp=sum_b_exp/(0.9*G)
```