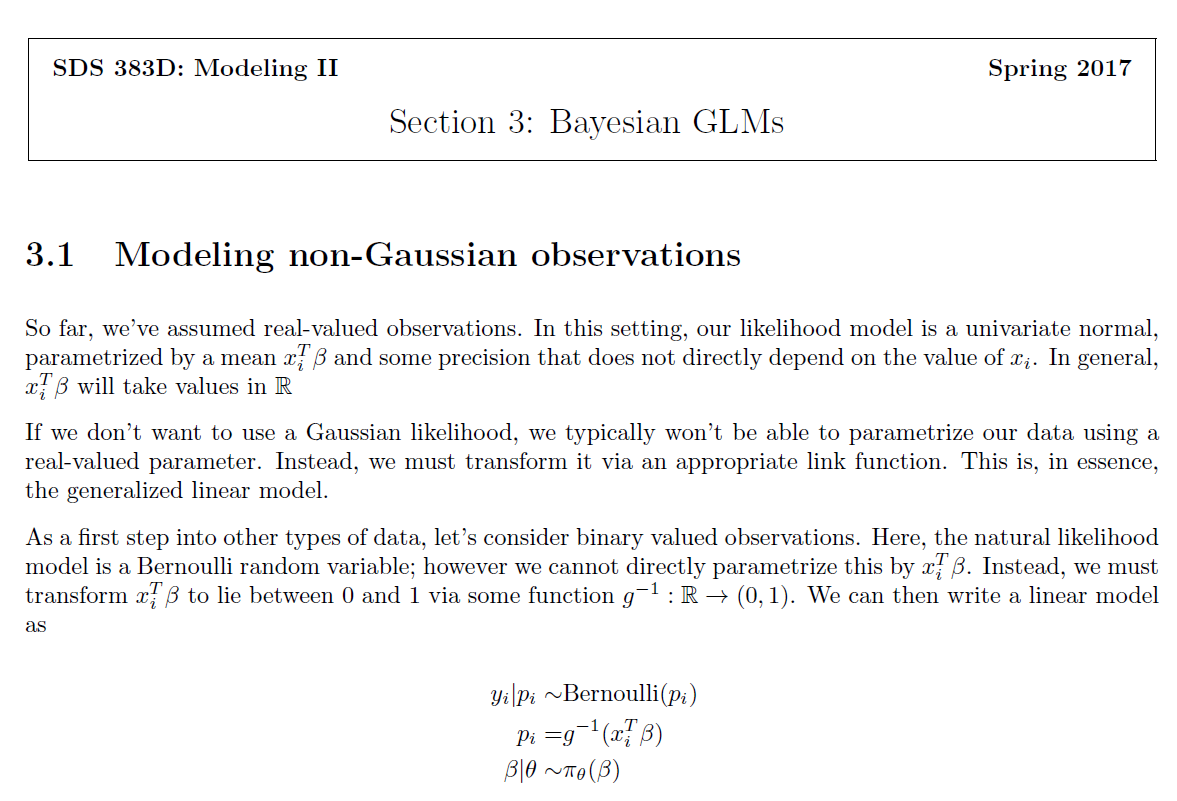
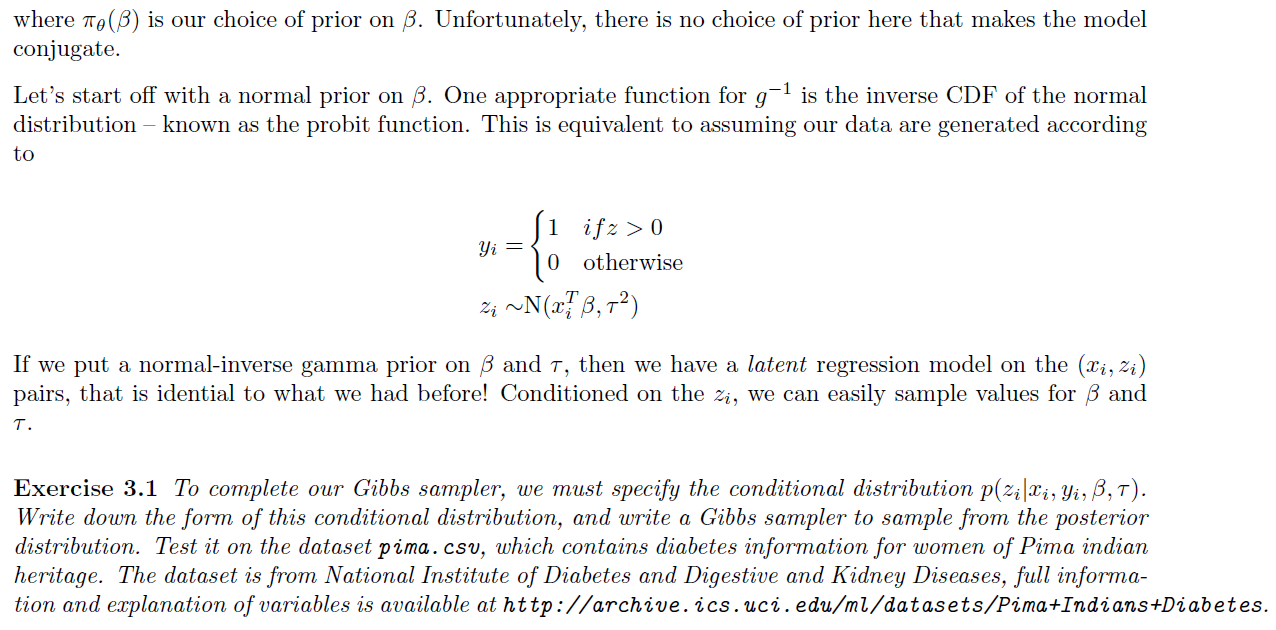
**Sareh Kouchaki**



\*\*\* useful link for generalized linear model: http://statmath.wu.ac.at/courses/heather\_turner/glmCourse\_001.pdf



Solution:

Useful Link: <https://rpubs.com/cakapourani/bayesian-binary-probit-model>

Performing inference for this model in the Bayesian framework is complicated by the fact that no conjugate prior π() exists. To overcome this problem, [Albert and Chib (1993)](http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476321) augmented the original model with an additional auxiliary variable that renders the conditional distributions of the model parameters equivalent to those under a Bayesian normal linear regression model with Gaussian noise; and derived an efficient Gibbs sampling scheme for computing the posterior statistics.

Augmented Model:

We are interested in computing the joint posterior distribution of the latent variables z and the model parameter given the data y and x.

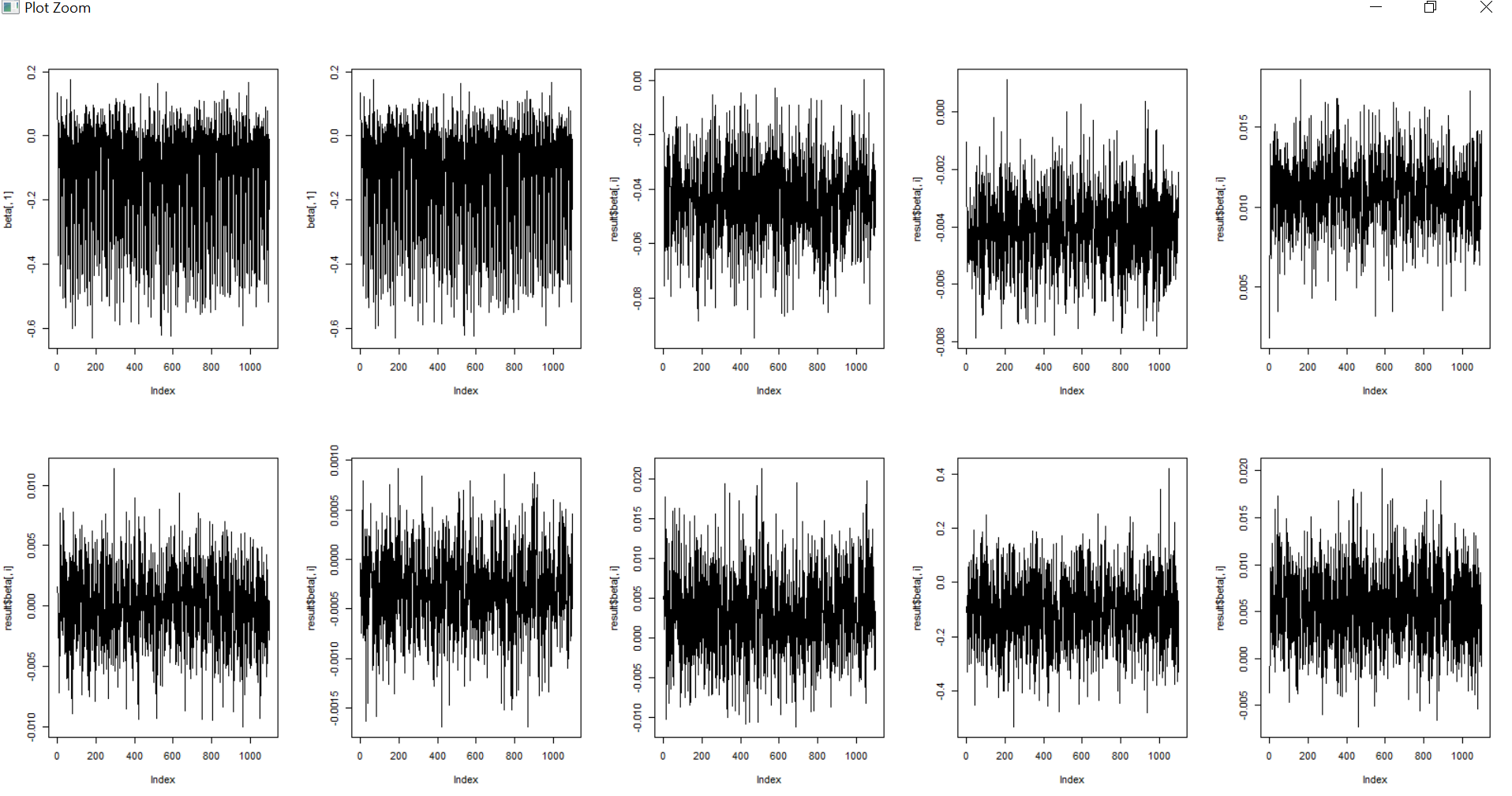
where we have,

where I is the indicator function, equal to 1 if the quantities inside the function are satisfied, and 0 otherwise.

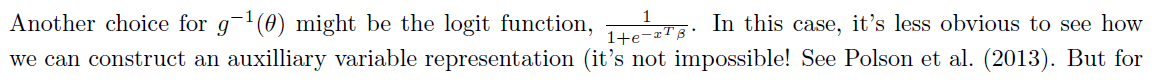
The joint posterior is difficult to normalize and sample from directly. However, computation of the marginal posterior of  and z using the Gibbs sampling requires only computing and .

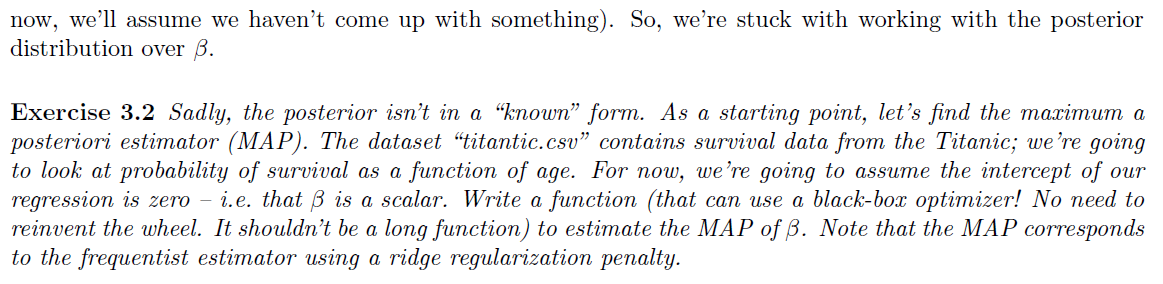
This is the posterior density for the normal linear regression. The posterior distribution would be normal.

If we assign a constant prior on , , then the conditional posterior would be as the following:



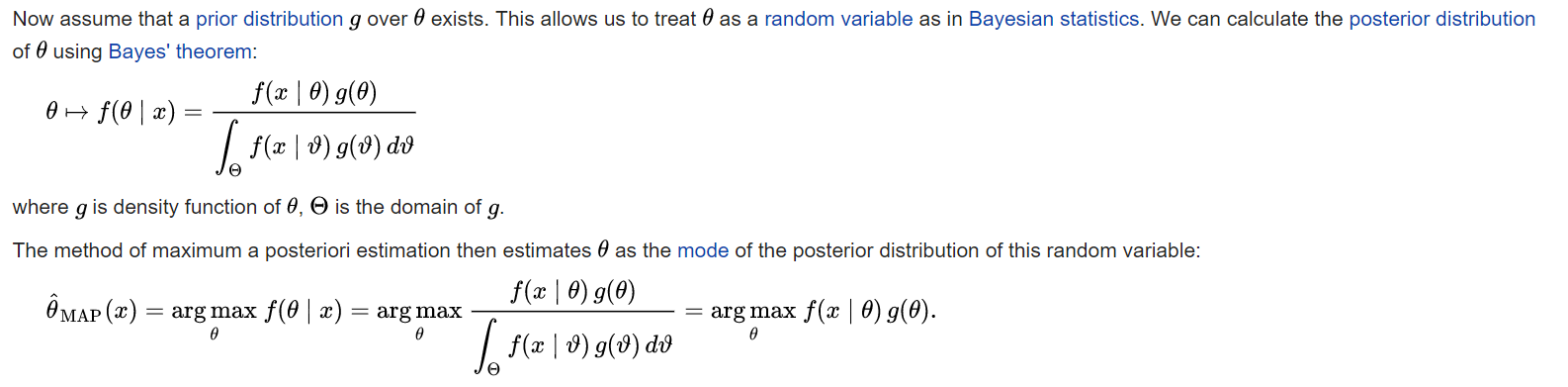
R code for Gibbs Sampling part is on the GitHub.





Solution:

Wikipedia: In Bayesian statistics, a maximum a posteriori probability (MAP) estimate is an estimate of an unknown quantity, that equals the mode of the posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data.

{\displaystyle {\hat {\theta }}\_{\mathrm {MAP} }(x)={\underset {\theta }{\operatorname {arg\,max} }}\ f(\theta \mid x)={\underset {\theta }{\operatorname {arg\,max} }}\ {\frac {f(x\mid \theta )\,g(\theta )}{\displaystyle \int \_{\Theta }f(x\mid \vartheta )\,g(\vartheta )\,d\vartheta }}={\underset {\theta }{\operatorname {arg\,max} }}\ f(x\mid \theta )\,g(\theta ).\!}

Logit regression:

Useful link: <https://www.cs.princeton.edu/~bee/courses/lec/lec_jan24.pdf>

For a Bernoulli distribution, with y {0,1} and p representing the probability of success, 0 p 1, we have:

Exponential family form of the Bernoulli distribution is as following:

In the Bernoulli distribution, in the exponential family, note that the logit function (i.e., log odds function) maps the mean parameter vector, p, to the natural parameter, η. The function that maps η to p is the logistic function, which is the inverse of the logit function as shown below:

logistic regression model:

As in linear regression, we have pairs of observed variables D = {(x1, y1), ..., (xn, yn)}.

* Observed input x is assumed to enter the model via a linear combination, .
* The conditional mean p is represented as a function of .
* The response y is characterized by an exponential family distribution with conditional mean p.

For logistic regression, we set our natural parameter Therefore, for our regression model where the conditional probability is modeled as a Bernoulli distribution, the parameter p = E[Y|X, can be obtained from the logistic function,

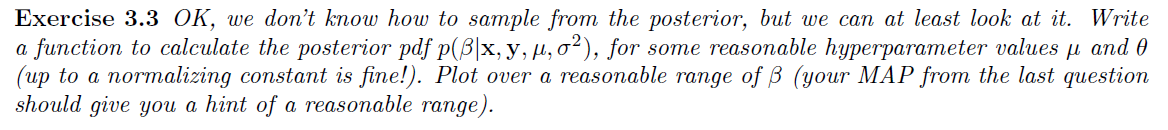
Likelihood function:

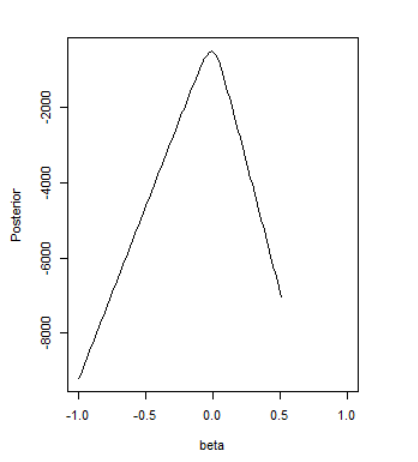
If we change p to , the likelihood function would be:

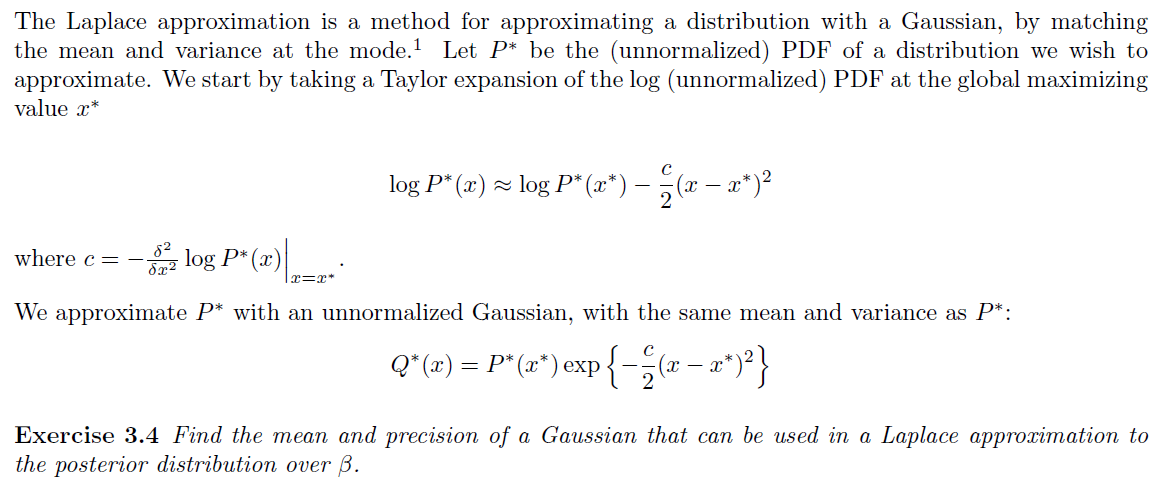
By considering Normal(0, I) as a prior on , the posterior distribution would be:

To find MAP, we need to maximize the above function.

The code is available on GitHub.







Answer:

