Chapter 1

Metropolis-Hastings Algorithm

For implementing Metropolis-Hastings algorithm simulated data were used with k=3 covariates and a constant term. A multivariate normal distribution was taken as a prior for β, $N(0, \Sigma)$ where Σ is a diagonal matrix with $1/\lambda$ on the diagonal. After considering a number of values for λ, $\lambda = 1$ was chosen.
The posterior distribution was simulated using as proposal a multivariate normal distribution centered at the current update of and with a covariance matrix given by the inverse of Fisher information evaluated at the current update.

Fisher information for $Y \sim Bernoulli(p)$ with probit link was derived based on the formula

$$I(\beta) = -E(\frac{\sigma^2 l}{\sigma\beta^2}) = \sum_{l}^{n} \frac{\phi^2(s_i)}{\Phi(s_i)(1 - \Phi(s_i))} x_i x_i' \tag{1.1}$$

where $s_i = \beta' x_i$

Two options were considered as a starting point for β: MLE estimates and least-square estimates. The latter produced better results.

Results:

The acceptance rate was 44.36%.

Table 1.1: Mean prediction for beta with $\lambda = 1$ and $\tau = 0.75$

β_i	β_i **mean prediction**	**95 % CI**	**True beta**
1	0.6773	(0.4435,0.9111)	0.8
2	0.6532	(0.416,0.8904)	0.7
3	0.476	(0.2579,0.6941)	0.4
3	0.4243	(0.1854,0.6632)	0.6

1.1 Fisher information

At each update of the algorithm, the proposal depends on the current position. In particular, the proposal is centered at the current position and has covariance matrix given by the inverse of the Fisher information.

In order to derive the Fisher information for a probit model, we can use the fact that

$$I(\beta) = -E(\frac{\sigma^2 l}{\sigma^2 \beta}) \tag{1.2}$$

Therefore, we proceed to compute the first derivative of the log-likelihood with respect to the coefficients, which equals

$$\frac{\sigma l}{\sigma \beta} = \sum_{y_i=1} \phi(s_i)\phi^{-1}(s_i)x_i - \sum_{y_i=0} \phi(s_i)(1-\phi(s_i))^{-1}x_i \tag{1.3}$$

where ϕ is the normal cumulative distribution function.

Differentiating one more time using the product rule, we find the second derivative, which equals

$$\begin{aligned} \frac{\sigma^2 l}{\sigma^2 \beta} = -\sum_{y_i=1} [\phi(s_i)\phi^{-1}(s_i) + \phi(s_i)\phi^{-2}(s_i)]x_i x_i\prime \\ + \sum_{y_i=0} [\phi^2(s_i)(1-\phi(s_i))^{-2} - s_i\phi(s_i)(1-\phi(s_i))]x_i x_i\prime \end{aligned} \tag{1.4}$$

Taking the expectation, we finally obtain

$$I(\beta) = -E(\frac{\sigma^2 l}{\sigma^2 \beta}) = \sum_{l}^{n} \frac{\phi^2(s_i)}{\phi(s_i)(1-\phi^2(s_i))} x_i x_i\prime \tag{1.5}$$

The Fisher Information matrix is a $K \times K$ matrix, where k is the number of explanatory variables in the model.

The following histograms show that betas converge to MLE estimates with β_1 and β_2 converging closer to true values then β_0 and β_3.

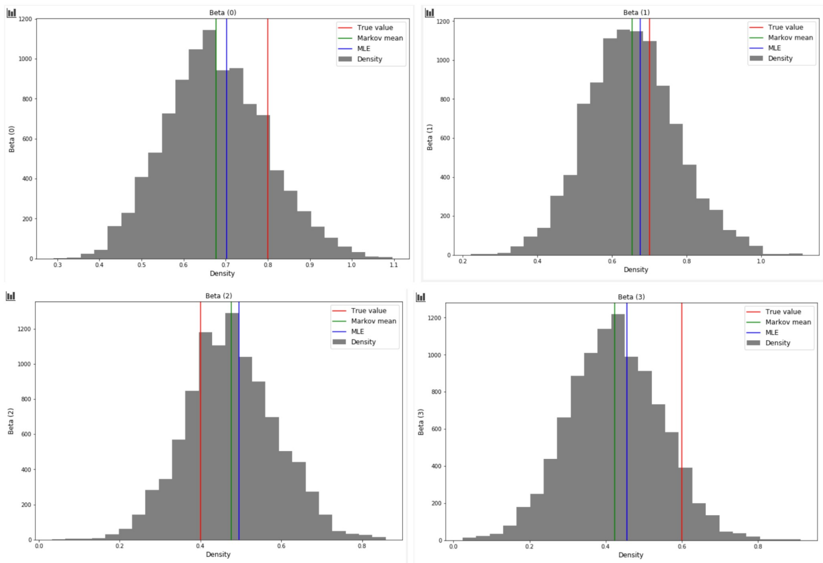

Figure 1.1: Histogram of the Markov chain with $\lambda = 1$, $\tau = 0.75$ and start points = MLE

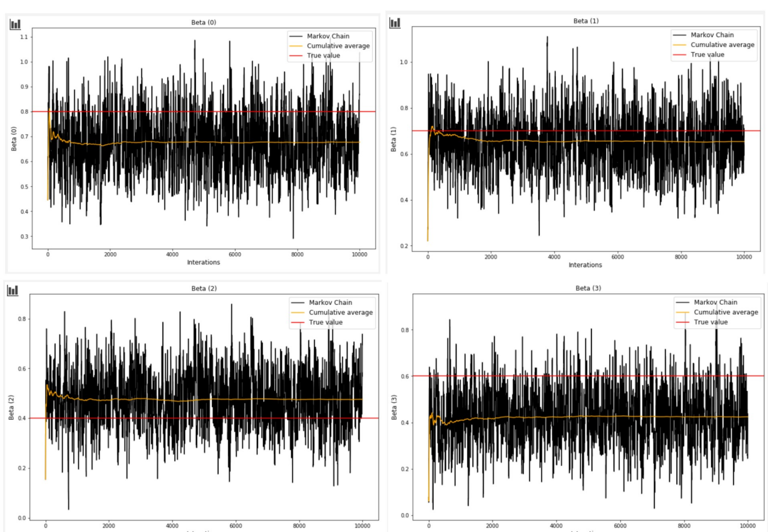

Figure 1.2: Trace plot of the markov chaing with $\lambda = 1$ and $\tau = 0.75$

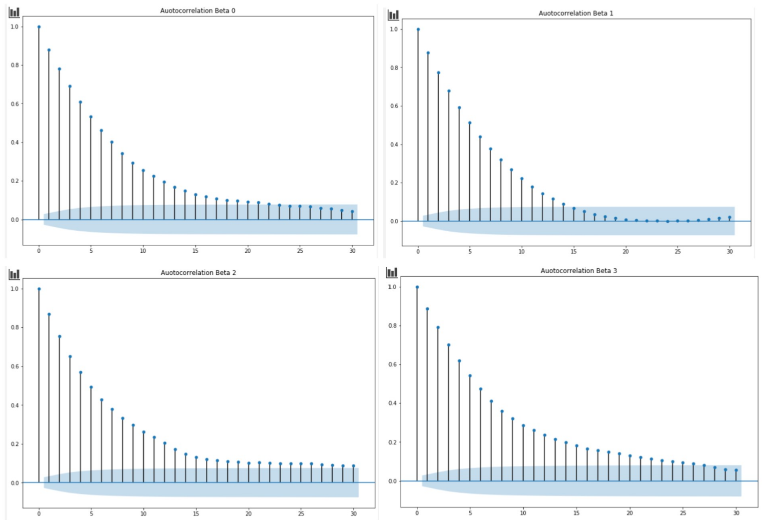

Figure 1.3: Autocorrelation graphic for the markov chain.

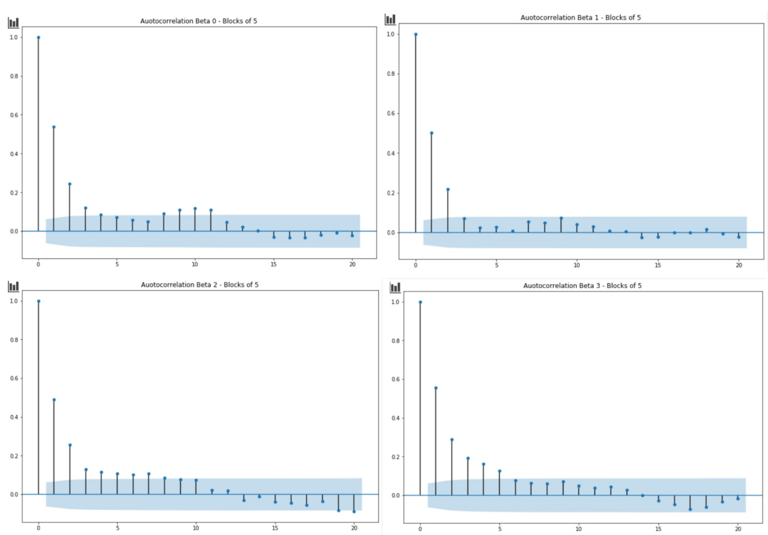

Figure 1.4: Autocorrelation graphic for the markov chain in steps of 5

Chapter 2

Gibbs sampler for binary data

We have N independent binary random variables where each y_i comes from a Bernoulli distribution with probability of success π_i.

We can complete the Bayesian model by taking a prior distribution $\pi(\beta)$ over the model parameters. Then the Bayesian binary probit regression model becomes,

$y_i \sim Bernoulli(\Phi(\eta_i))$ where $\eta_i = x_i\beta$ The posterior distribution for the Bayesian binary probit model is the following,

$$p(\beta|y,X) \propto p(\beta)p(y|\beta,X) = \pi(\beta)\prod_{i=1}^{N} p(y_i|\beta,x_i) = \pi(\beta)\prod_{i=1}^{N} \Phi(x_i\beta)^{y_i}(1-\Phi(x_i\beta))^{1-y_i} \tag{2.1}$$

Performing inference for this model in the Bayesian framework is complicated by the fact that no conjugate prior $\pi(\beta)$ exists for the parameters of the probit regression model. Based on an article by Albert and Chib (1993) , we have considered a multivariate normal distribution.

Let $z_i, \ldots, z_n$, be n auxiliary variables defined as follows.

$$y_i = \begin{cases} 1 & if z_i > 0 \\ 0 & if z_i < 0 \end{cases} \tag{2.2}$$

$Z_1, Z_2, \ldots, Z_n$ are independent variables with normal density having expected value $x_i^T\beta$ and variance 1

$\pi(\beta|Z,Y)$ the posterior distribution of the vector of parameters of a multivariate linear regression model is as follows

$$\pi(\beta|y,Z,X) \propto p(\beta)p(z|\beta,X) = \pi(\beta)\prod_{i=1}^{N} p(z_i|\beta,x_i) \tag{2.3}$$

where we have,

$$p(z_i|\beta, x_i) = N(z_i|x_i\beta, 1)p(y_i|z_i) = 1(y_i = 1)1(z_i > 0) + 1(y_i = 0)1(z_i <= 0) \tag{2.4}$$

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

This joint posterior is difficult to normalize and sample from directly. However, computation of the marginal posterior of β and z using the Gibbs sampling requires only computing $p(\beta|z, y, X)$ and $p(z|\beta, y, X)$, and these full conditional distributions are of standard forms. It should be noted that $p(\beta|z, y, X) = p(\beta|z, X)$, since β is conditionally independent of y given z.

First, the full conditional of β is given by,

$$\pi(\beta|Z, X) \sim \pi(\beta) \prod_{i=1}^{N} N(z_i|x_i\beta, 1) \tag{2.5}$$

This quantity is the posterior density for the normal linear regression model. Using standard linear model results, if we use a constant prior for β, i.e. $\pi(\beta) \propto 1$, then,

$$\pi(\beta|Z, X) \propto N((X^T X)^{-1} X^T z, (X^T X)^{-1}) \tag{2.6}$$

now for full conditional of $\pi(Z|\beta, Y)$ we have:

$z_i \sim N(x_i\beta, 1)$, Thus, the full conditional of $z_i|\theta, y_i, x_i$is a truncated normal distribution,

$$z_i|\theta, y_i, x_i = \begin{cases} TN(x_i\beta, 1, 0, \infty) & if y_i = 1 \\ TN(x_i\beta, 1, -\infty, 0) & if y_i = 0 \end{cases} \tag{2.7}$$

The results of implementing Auxiliary variables Gibbs Sampler are the following:

β_i	β_i **mean prediction**	**95 % CI**	**True beta**
1	0.7186	(0.4769,0.9603)	0.8
2	0.6914	(0.4589,0.9239)	0.7
3	0.5046	(0.2823,0.7269)	0.4
4	0.4666	(0.2198,0.7134)	0.6

Compared to MH, Gibbs sampling algorithm performs better in terms of convergence to true values. However, it appeared to be less efficient in terms of running time.

The histograms show that betas converge to MLE estimates with β_0 and β_1 converging closer to true values then β_2 and β_3. The non informative Gibbs sampler gave us a worse result.

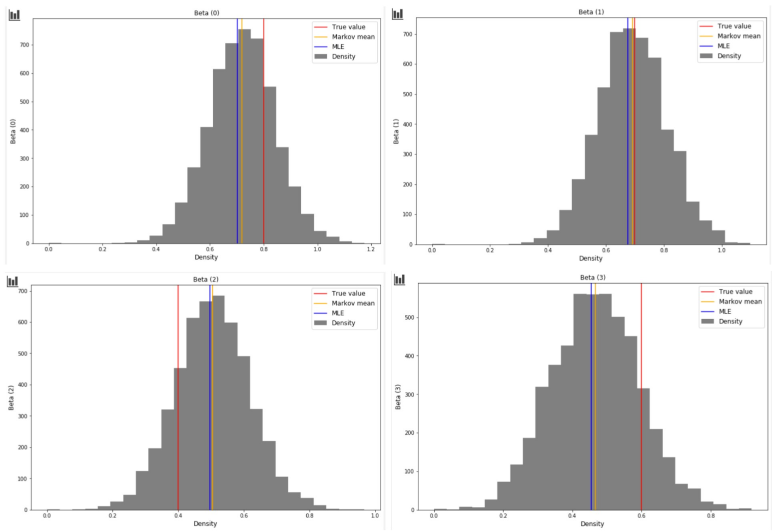

Figure 2.1: Histogram of the informative Gibbs sampler with λ = 1 and start points = MLE

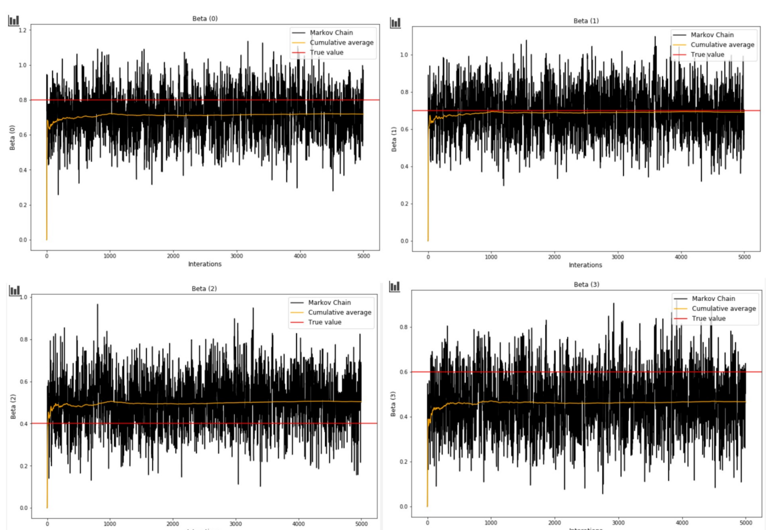

Figure 2.2: Traceplots of the informative Gibbs sampler markov chain with λ = 1 and start points = MLE

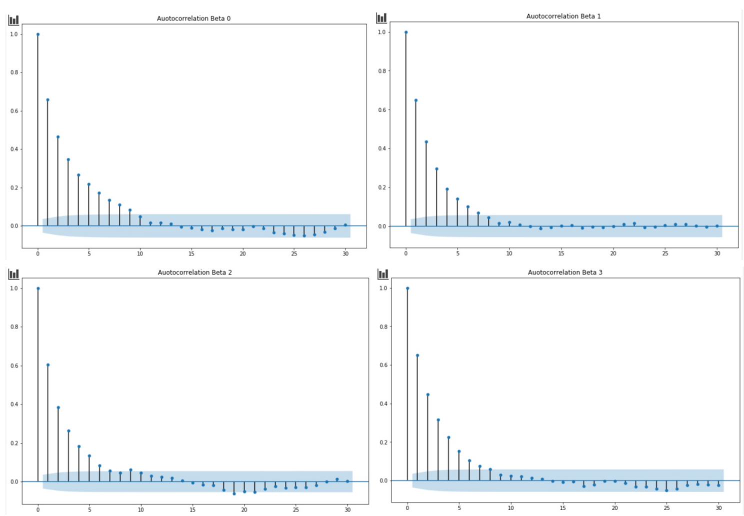

Figure 2.3: Autocorrelation graphic for the markov chain Gibbs Sampler markov chain.

Appendices

τ	0.5	0.75	1.00	2.00	3.00
Acceptance Rate	50.48%	44.36%	37.12%	22.60%	15.55%

Figure 2.4: Metropolis-Hasting Algorithm: Different values of tau and its acceptance rate for the Markov, for $\lambda = 1$

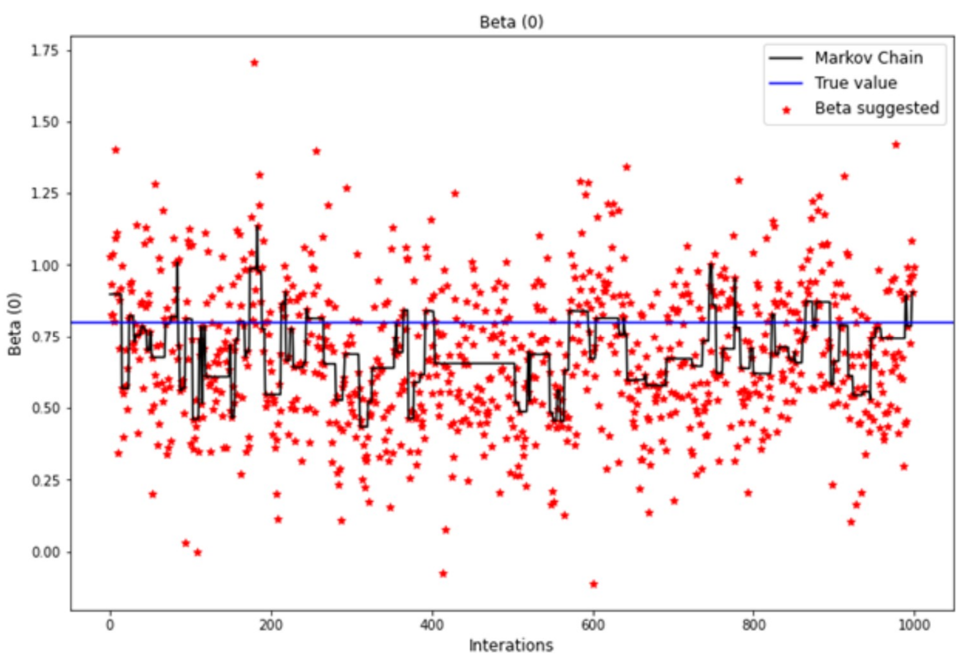

Figure 2.5: Traceplot for block of 1.000 iterations of the algorithm with $\tau = 3$

Bibliography

[1] Demidenko, E. (2001). Computational aspects of probit model. Mathematical Communications, 6, 233-247.

[2] Albert, J. H., and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association, 88(422), 669-679.