Homework 1

Ian Frankenburg

4/7/2020

Generalized Linear Models

Part (1).

Describe and implement a Metropolis-Hastings algorithm designed to obtain a MC with stationary distribution $p(\beta|Y)$

First I want to think about the likelihood times prior: $\prod p(y_i|X_i,\beta) \times p(\beta|X_i)$. This is theoretically how I'd do it, but a major conceptual point here is that it's going to be hard to generate a sensible multivariate proposal distribution. Instead, I'm gonna sample component-wise ie sample each β_j one at a time. The downside of this is mixing rate: Imagine I could sample a sub-vector of $\boldsymbol{\beta}$. Then I could take advantage of possible correlation to make better proposals and speed up convergence. This will give me a posterior target, afterwhich I can implement an MCMC-type algirithm

```
logtarget <- function(y,X, beta){
  n <- nrow(X)
  likelihood <- p <- 0
  for(i in 1:n){
    p <- pnorm(t(X[i,])%*%beta)
    likeihood <- log(p)^y[i] + log(1-p)^(1-y[i]) + likeihood
  }
  return(likelihood - 1/(2*n) *t(X%*%beta)%*%(X%*%beta))
}
fit <- summary(glm(y~X,family=binomial(link="probit")))
beta <- fit$coeff[,1]
for(s in 1:samples){
  # propose new beta based on symmetric distribution

  # evaluate step in MH

}
```

Part (2).

Describe and implement a data augmented (DA-MCMC) strategy targeting $p(\beta|Y)$

The original model is

$$P(y_i = 1|x_i, \beta) = \Phi(x_i^\top \beta).$$

This is equivalent to the model $P(y_i = 1|x_i, \beta) = P(\mathbf{1}(z_i > 0) = 1) = P(z_i > 0)$, where $z_i \overset{\text{iid}}{\sim} N(x_i^\top \beta, 1)$. This follows immediately since

$$\int_{-\infty}^{x_i^\top \beta} N(t;\ 0,1)dt = \int_{-\infty}^{0} N(z_i; x_i^\top \beta, 1)dz_i$$

by a change of variables $z_i := t - x_i^\top \beta$. Thus $\Phi(x_i^\top \beta) = P(z_i > 0)$.

In defining the latent model, the full conditionals $p(\beta|\boldsymbol{y}, z)$, $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta})$ become tractible, so I can use a Gibbs sampler.

I'll start with $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{z})$

$$\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{z}) &\propto p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\beta})p(\boldsymbol{\beta}) \\
&= \cancel{p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{z})}p(\boldsymbol{z}|\boldsymbol{\beta})p(\boldsymbol{\beta}) = p(\boldsymbol{z}|\boldsymbol{\beta})p(\boldsymbol{\beta}) \\
&= N(\boldsymbol{z}; X\boldsymbol{\beta}, \boldsymbol{I})N(\boldsymbol{\beta}; \mathbf{0}, n(X^\top X)^{-1})
\end{aligned}$$

Now for $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta})$.

$$\begin{aligned}
p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}) &\propto p(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{z}) \\
&\propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{z})p(\boldsymbol{z}|\boldsymbol{\beta})\cancel{p(\boldsymbol{\beta})} \\
&\propto \prod_{i=1}^{n} [\mathbf{1}(y_i = 1)\mathbf{1}(z_i > 0) + \mathbf{1}(y_i = 0)\mathbf{1}(z_i < 0)]N(\boldsymbol{z}; X\boldsymbol{\beta}, \boldsymbol{I})
\end{aligned}$$

Since our sampling model assumes the y's are independent, so are the z's and I can sample the full conditionals independently, i.e.

$$\begin{aligned}
p(z_i|\boldsymbol{y}, \boldsymbol{\beta}) &\propto p(\boldsymbol{y}, \boldsymbol{\beta}, z_i) \propto p(y|\boldsymbol{\beta}, z_i)p(z_i|\beta) \\
&\propto 1(y_i = 1)\mathbf{1}(z_i > 0) + \mathbf{1}(y_i = 0)\mathbf{1}(z_i < 0)]N(z_i; x_i^\top \boldsymbol{\beta}, 1) \\
&= \begin{cases} N(z_i; x_i^\top \boldsymbol{\beta}, 1) * \mathbf{1}_{[0,\infty)}(z_i) & \text{if } y_i = 1 \\ N(z_i; x_i^\top \boldsymbol{\beta}, 1) * \mathbf{1}_{(-\infty,0)}(z_i) & \text{if } y_i = 0 \end{cases}
\end{aligned}$$

Now I can implement a Gibbs sampler to iteratively draw from these conditionals.

Result: Samples from joint posterior $p(\boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y})$

for *s in # samples* **do**

end

Part (5).

For logit model, describe and implement a data augmented (DA-MCMC) strategy targeting $p(\beta|Y)$

Similar to the probit case, the posterior is intractible but can be written as

$$\begin{aligned}
p(\beta|\boldsymbol{y}) &\propto \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} p(\beta) \\
&= \prod_{i=1}^{n} \left[\frac{e^{x_i^\top \beta}}{1+e^{x_i^\top \beta}}\right]^{y_i} \left[1 - \frac{e^{x_i^\top \beta}}{1+e^{x_i^\top \beta}}\right]^{1-y_i} N(\boldsymbol{\beta}; 0, n(X^\top X)^{-1}) \\
&= \prod_{i=1}^{n} \frac{(e^{x_i^\top \beta})^{y_i}}{1+e^{x_i^\top \beta}} N(\boldsymbol{\beta}; 0, n(X^\top X)^{-1})
\end{aligned}$$

This will be my target in the random walk MH. Therefore, an algorithm might be

Result: Samples from joint posterior $p(\boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y})$

for *s in # samples* **do**

  $\ell(\beta) := \log L(\beta) = \sum_{i=1}^{n} y_i x_i^\top \beta - \log(1+e^{x_i^\top \beta}) - \frac{1}{2n}\beta^\top (X^\top X)\beta$

end

Part (6).

Describe and implement a Langevin-Hastings algorithm designed to obtain a MC with stationary distribution $p(\beta|Y)$

The Langevin-Hastings algorithm will work by utilizing the a 2^{nd} order Taylor approximation of the target distribution. I'm going to use a component-wise MCMC algorithm, so I'm going to make proposals for each β_j Therefore, my proposal for β_j at iteration s will be something like

$$\beta_j^* \sim N\Big(\beta_j^{(s)} + \frac{1}{2}\sigma^2 \frac{\partial}{\partial \beta_j^{(s)}} \log\big(p(\boldsymbol{y}|\beta_j)p(\boldsymbol{\beta})\big), \sigma^2 \boldsymbol{I}\Big)$$

Earlier I showed the poserior is proportional to

$$\prod_{i=1}^{n} \frac{(e^{x_i^\top \beta})^{y_i}}{1 + e^{x_i^\top \beta}} N(\boldsymbol{\beta}; 0, n(X^\top X)^{-1})$$

so the log-posterior is

$$\begin{aligned}
&\sum_{i=1}^{n} y_i x_i^\top \boldsymbol{\beta} - \log(1 + e^{x_i^\top \boldsymbol{\beta}}) - \frac{1}{2n}\beta^\top (X^\top X)\beta \\
\Rightarrow &\nabla\Big\{\sum_{i=1}^{n} y_i x_i^\top \boldsymbol{\beta} - \log(1 + e^{x_i^\top \boldsymbol{\beta}}) - \frac{1}{2n}\boldsymbol{\beta}^\top (X^\top X)\beta\Big\} \\
= &\sum_{i=1}^{n} [y_i - \frac{e^{x_i^\top \boldsymbol{\beta}}}{1 + e^{x_i^\top \boldsymbol{\beta}}}] x_i^\top - \frac{1}{n}\boldsymbol{\beta}^\top (X^\top X)
\end{aligned}$$

Therefore, the proposal will be of the form

$$\beta_j^* \sim N(\beta_j + \frac{1}{2}\sigma^2 \frac{\partial}{\partial \beta_j}\text{log-target}, \sigma^2)$$

Where $\frac{\partial}{\partial \beta_j}$log-target is the j-th element of the gradient ∇ computed previously.

Part (7).

Describe and implement an adaptive Metropolis-Hastings algorithm designed to obtain a MC with stationary distribution $p(\beta|Y)$