

MATH 598 - PROJECT 2

Please submit your project by 11.59 pm on Wednesday 13th March by uploading a single pdf to myCourses.

1. Use **rejection sampling** to sample directly from the posterior distribution for the two parameters (β_0, β_1) in the binomial regression problem where

$$Y_i | x_i, n_i \sim \text{Binomial}(n_i, p(x_i, \beta)) \quad i = 1, \dots, K$$

and where

$$\log \left(\frac{p(x, \beta)}{1 - p(x, \beta)} \right) = \beta_0 + \beta_1 x$$

for x a continuous (predictor) variable, where the data are given as follows: These data record the

i	1	2	3	4	5	6	7	8
n_i	20	23	23	23	17	24	19	18
y_i	16	17	20	19	4	7	0	3
x_i	0.32	1.25	1.62	2.54	4.95	6.46	8.85	8.98

number of patients y_i (out of n_i) who failed to react to a pain relief treatment when the dose level was x_i , for $i = 1, \dots, K$, with $K = 8$.

Assume that the prior distribution, $\pi_0(\beta)$ is a bivariate Normal distribution where β_0 and β_1 are presumed to be independent, zero mean and to have prior variance 25.

Your objective is to make the rejection sampling procedure as efficient as possible, that is, to have the highest possible acceptance rate for generated points. 20 Marks

2. **Sampling Importance Resampling (SIR)** is a technique that appeals to importance sampling principles to produce an approximate sample from a density $f(y)$ using samples from an importance density $f_0(y)$ as follows:

- (i) Sample $y_1, \dots, y_N$ from $f_0(y)$.
- (ii) Form sampling weights

$$w_i = \frac{f(y_i)/f_0(y_i)}{\sum_{j=1}^N (f(y_j)/f_0(y_j))} \quad i = 1, \dots, N$$

- (iii) Resample (with replacement) n points $\{y_1^*, \dots, y_n^*\}$ from the set $\{y_1, \dots, y_N\}$ with associated sampling weights $\{w_1, \dots, w_N\}$.

The sample $\{y_1^*, \dots, y_n^*\}$ is an approximate sample from f . Some example code can be found on the next page.

Use SIR to produce an approximate sample from the posterior distribution in Question 1.

10 Marks

Sampling Importance Resampling Code for resampling $n = 2000$ points from $N = 100000$ for $f(y) \equiv \text{Normal}(2, 1)$ using $f_0(y) \equiv \text{Student}(5)$.

```

1 N<-100000
2 Y<-rt(N, 5)                #Sample
3 w<-dnorm(Y, 2, 1)/dt(Y, 5)  #Form the weights
4 w<-w/sum(w)
5
6 n<-2000
7 Yst<-sample(Y, prob=w, rep=T, size=n) #Perform the resampling.
8
9 hist(Yst[Yst>-6 & Yst<6], breaks=seq(-6, 6, by=0.2), col='gray',
10  xlab='y', ylim=range(0, 0.5), prob=T, main='SIR results')
11 box()
12 xv<-seq(-6, 6, length=1001)
13 yv<-dnorm(xv, 2, 1)
14 lines(xv, yv, col='red', lwd=2)
15 tv<-dt(xv, 5)
16 lines(xv, tv, col='blue', lwd=2)
17 legend(-6, 0.5, c('f0', 'f'), lty=1, lwd=2, col=c('blue', 'red'))

```

Here is the resulting sample:

