Homework 6

Kalvin Goode, Amil Khan

December 8, 2018

Logistic Regression for Toxicity Data

A pest control company is testing the efficacy of a poison by exposing adult flour beetles to a gaseous compound. Their experiments proceed by administering various dose levels of the poison to to batches of beetles. The beetles' responses are typically characterized by a binary outcome (e.g. dead or alive). An experiment of this kind gives rise to data where each observation is a triplet (x_i, n_i, y_i). x_i represents the log dosage in group i given to n_i beetles, of which y_i end up dying as a result of exposure. The company runs 4 experiments, each at a different dosages, on 5 beetles each. The resulting data can be seen below:

Part a.)

Solve for $\theta_i(x_i)$ as a function of α and β by inverting the logit function. Use this to write out the joint sampling distribution, $\prod_{i=1}^{N} p(y_i \mid \alpha, \beta, x_i, n_i)$.

Solution.

Assume that the number of beetle deaths Y_i given a chemical dosage, x_i, is

$$Y_i \sim \text{Binomial}(\theta(x_i), n_i)$$

where $\theta(x_i)$ is the probability of death given dosage x_i. We will assume that $\text{logit}(\theta_i(x_i)) = \alpha + \beta x_i$, where $\text{logit}(\theta)$ is defined as $\log\left(\frac{\theta}{1-\theta}\right)$. Hence substituting in what we know,

$$\begin{aligned}
\text{logit}(\theta_i(x_i)) &= \alpha + \beta x_i \\
\log\left(\frac{\theta_i(x_i)}{1-\theta_i(x_i)}\right) &= \alpha + \beta x_i
\end{aligned}$$

To get rid of the log, we exponentiate both sides and multiply both sides by $(1 - \theta_i(x_i))$, which gives us

$$\begin{aligned}
\exp^{\log\left(\frac{\theta_i(x_i)}{1-\theta_i(x_i)}\right)} &= \exp^{\alpha+\beta x_i} \\
\theta_i(x_i) &= \exp^{\alpha+\beta x_i}(1 - \theta_i(x_i))
\end{aligned}$$

Next, we simply distribute and factor both sides to solve for $\theta_i(x_i)$, which gives us,

$$\begin{aligned}
\theta_i(x_i)(1 + \exp^{\alpha+\beta x_i}) &= \exp^{\alpha+\beta x_i} \\
\theta_i(x_i) &= \frac{\exp^{\alpha+\beta x_i}}{1 + \exp^{\alpha+\beta x_i}}
\end{aligned}$$

Finally we will write out the joint sampling distribution, $\prod_{i=1}^{N} p(y_i \mid \alpha, \beta, x_, n_i)$, and simplify,

$$
\begin{aligned}
&\prod_{i=1}^{N}\binom{n_i}{y_i}(\theta_i(x_i))^{y_i}(1-\theta_i(x_i))^{n_i-y_i} \\
&\propto \prod_{i=1}^{N}\binom{n_i}{y_i}\left(\frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}}\right)^{y_i}\left(1-\frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}}\right)^{n_i-y_i} \\
&\propto \left(\frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}}\right)^{\sum y_i}\left(1-\frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}}\right)^{\sum n_i-y_i}
\end{aligned}
$$

Part b.)

The dose at which there is a 50% chance of being lethal, $\theta(x_i) = 0.5$ is known as LD50, and is often of interest in toxicology studies of this type. Solve for LD50 as a function of α and β.

Solution.

For $\theta(x_i) = 0.5$, we have,

$$
\begin{aligned}
\prod_{i=1}^{N}(\theta_i(x_i))^{y_i}(1-\theta_i(x_i))^{n_i-y_i} \implies & \prod_{i=1}^{N}(0.5)^{y_i}(0.5)^{n_i-y_i} \\
& \propto (0.5)^{\sum y_i}(0.5)^{\sum n_i-y_i}
\end{aligned}
$$

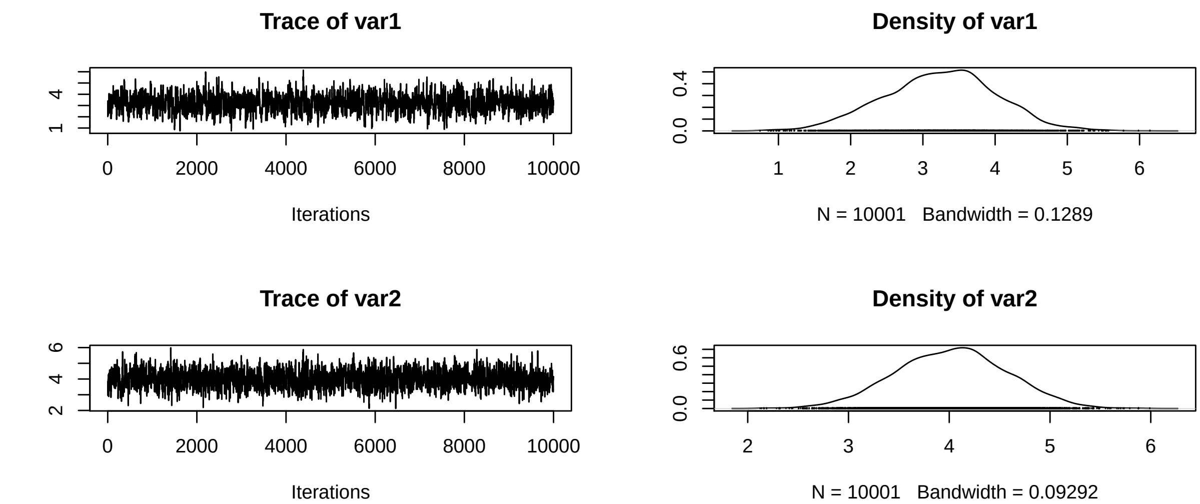
So, using what we found earlier,

$$
\begin{aligned}
\theta_i(x_i) &= \frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}} \\
0.5 &= \frac{\exp^{\alpha+\beta x_i}}{1+\exp^{\alpha+\beta x_i}} \\
\exp^{\alpha+\beta x_i} &= \frac{0.5}{1-0.5}
\end{aligned}
$$

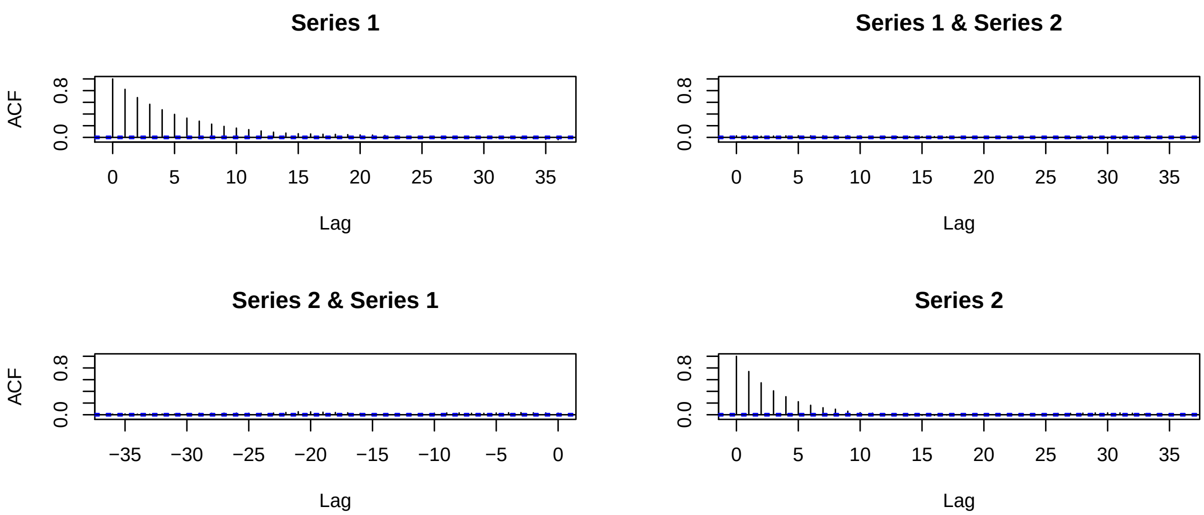
Taking the log of both sides and solving for x_i,

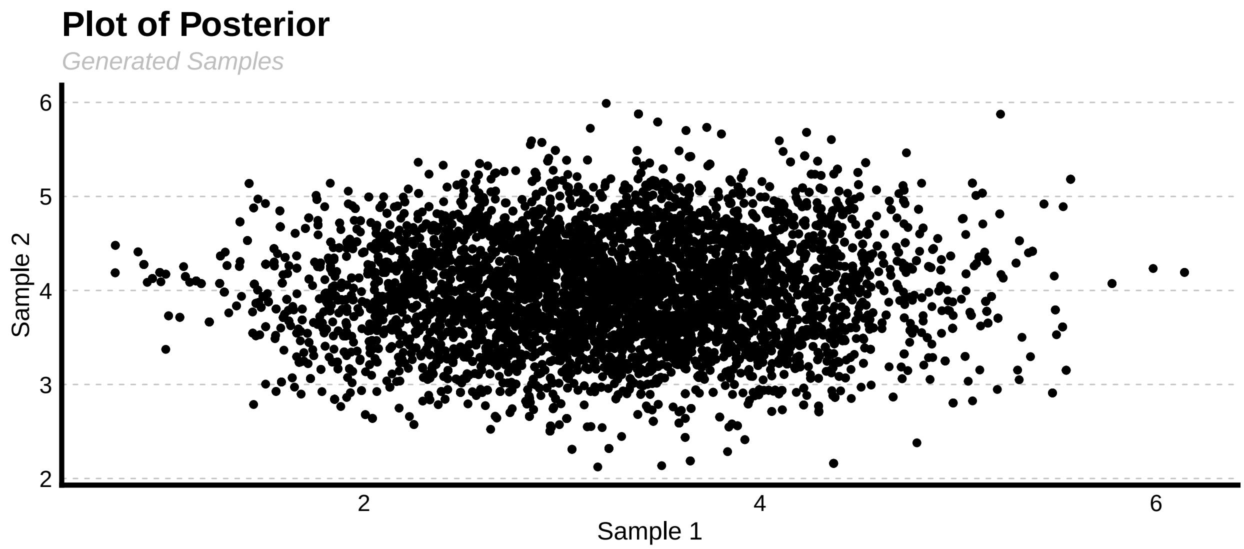
$$
\begin{aligned}
\alpha+\beta x_i &= \log\left(\frac{0.5}{1-0.5}\right) \\
\alpha &= -\beta x_i \\
x_i &= -\frac{\alpha}{\beta}
\end{aligned}
$$

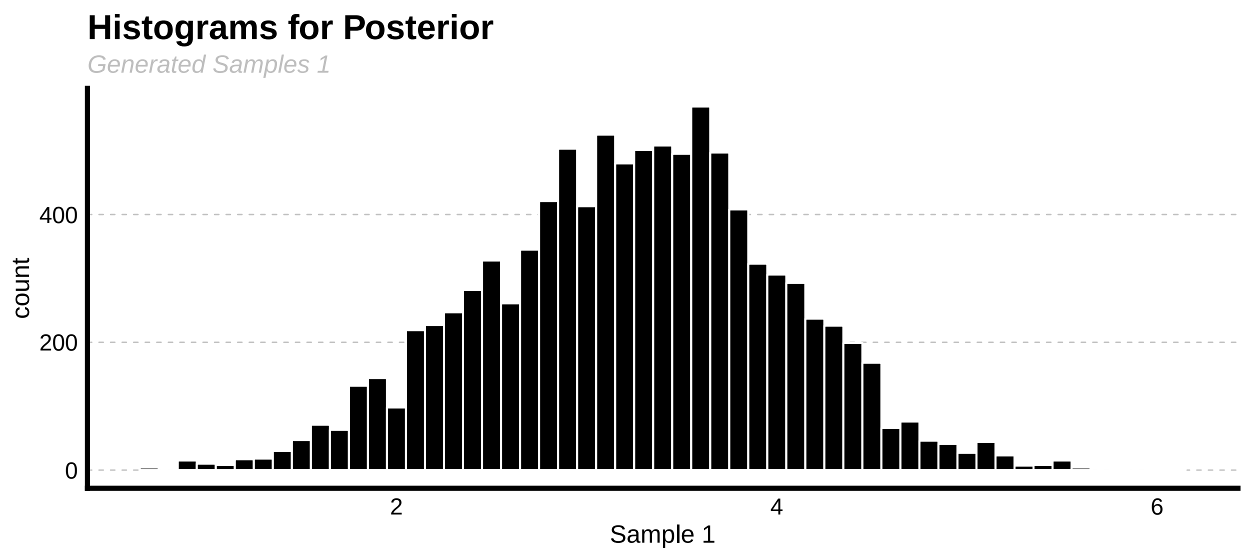
Part c.)

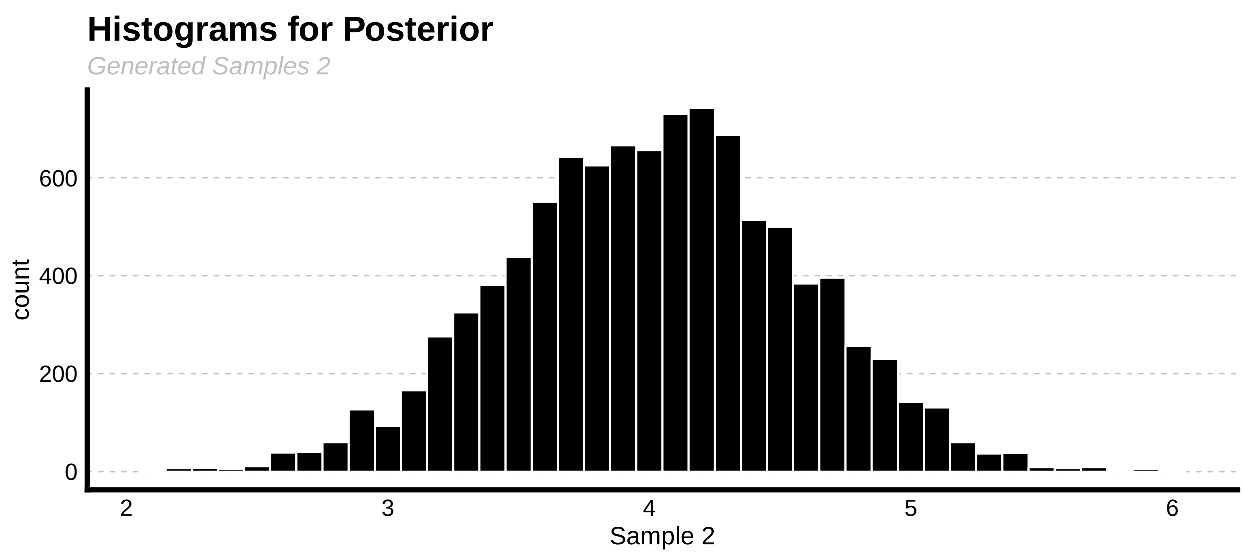



```
##      var1      var2
##  946.0696 1573.8476
```




Plot of Posterior
Generated Samples
Sample 2
6
5
4
3
2
2
4
6
Sample 1


Histograms for Posterior
Generated Samples 1
count
400
200
0
2
4
6
Sample 1



Part d.)

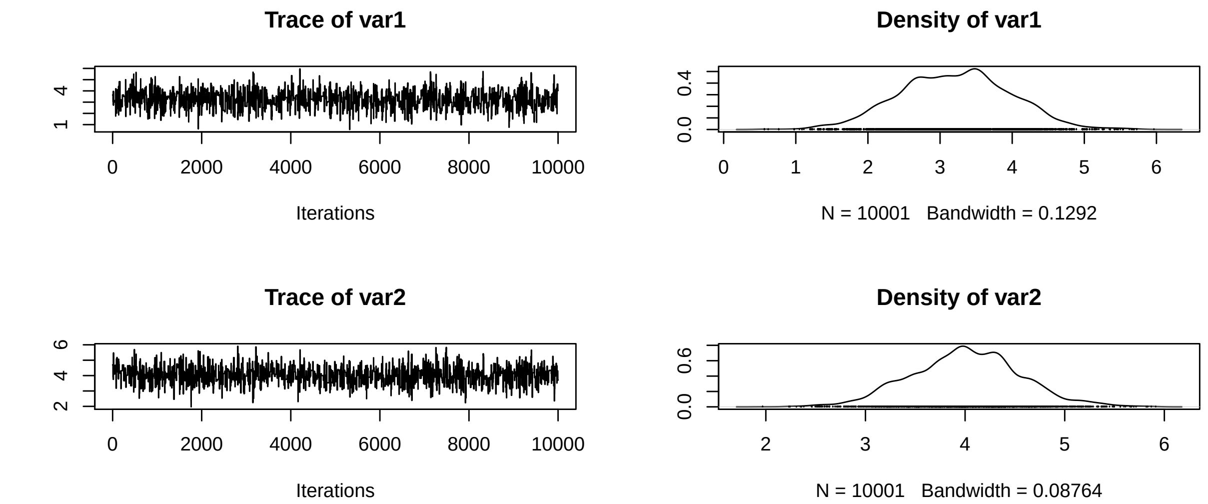
```
xgrid <- seq(-1, 1, by = 0.01)
samples <- rw_metrop_multi(c(10, 10), 1000, 10000, matrix(c(5, 2.5, 2.5, 5), nrow = 2))
length(samples)
```

```
## [1] 20002
```

```
compute_curve <- function(samp) {
beta_0 <- samp[1]
beta_1 <- samp[2]
beta_0 + beta_1 * xgrid
}
res <- apply(samples, 1, compute_curve)
quantiles <- apply(res, 1, function(x) quantile(x, c(0.025, 0.25, 0.75, 0.975)))
posterior_mean <- rowMeans(res)
effectiveSize(samples)
```

```
##      var1     var2
## 831.9741 931.1507
```

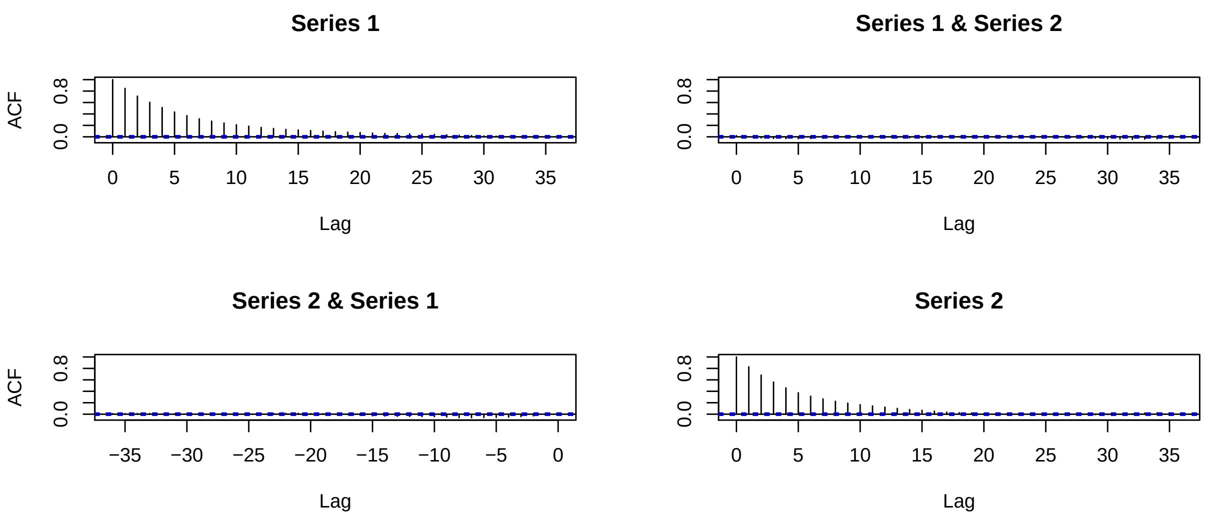
```
plot(as.mcmc(samples))
```



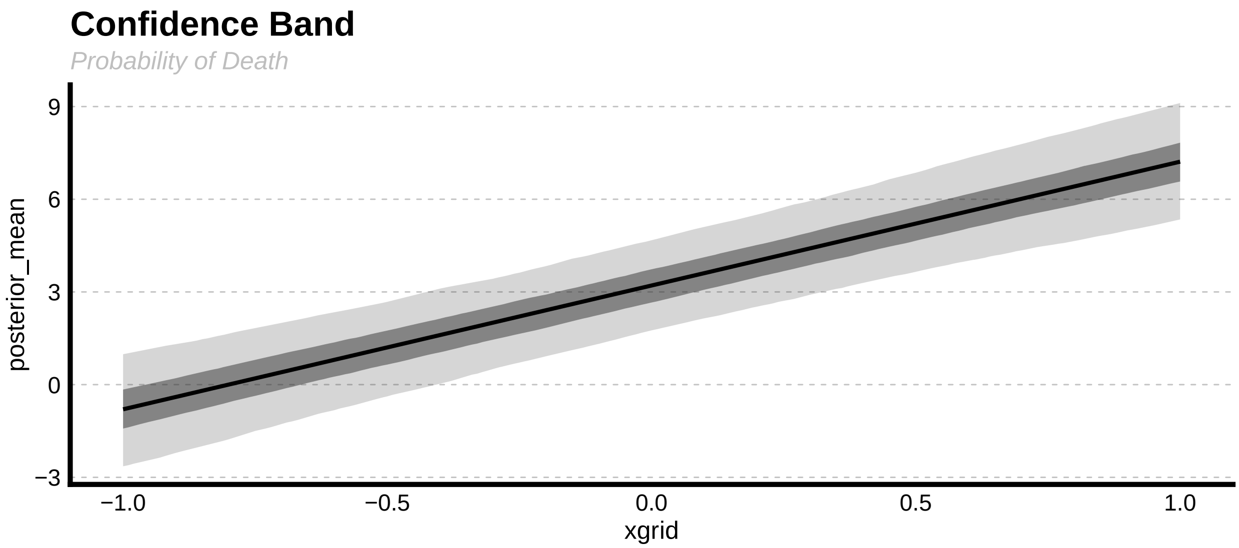

```
effectiveSize(samples)
```

```
##      var1     var2
## 831.9741 931.1507
```

```
acf(samples)
```



Part e.)


Confidence Band
Probability of Death
posterior_mean
9
6
3
0
−3
−1.0
−0.5
0.0
0.5
1.0
xgrid