STAT 578 - Advanced Bayesian Modeling - Fall 2019
Assignment 4

Xiaoming Ji

Solution for Problem (a)

(i)

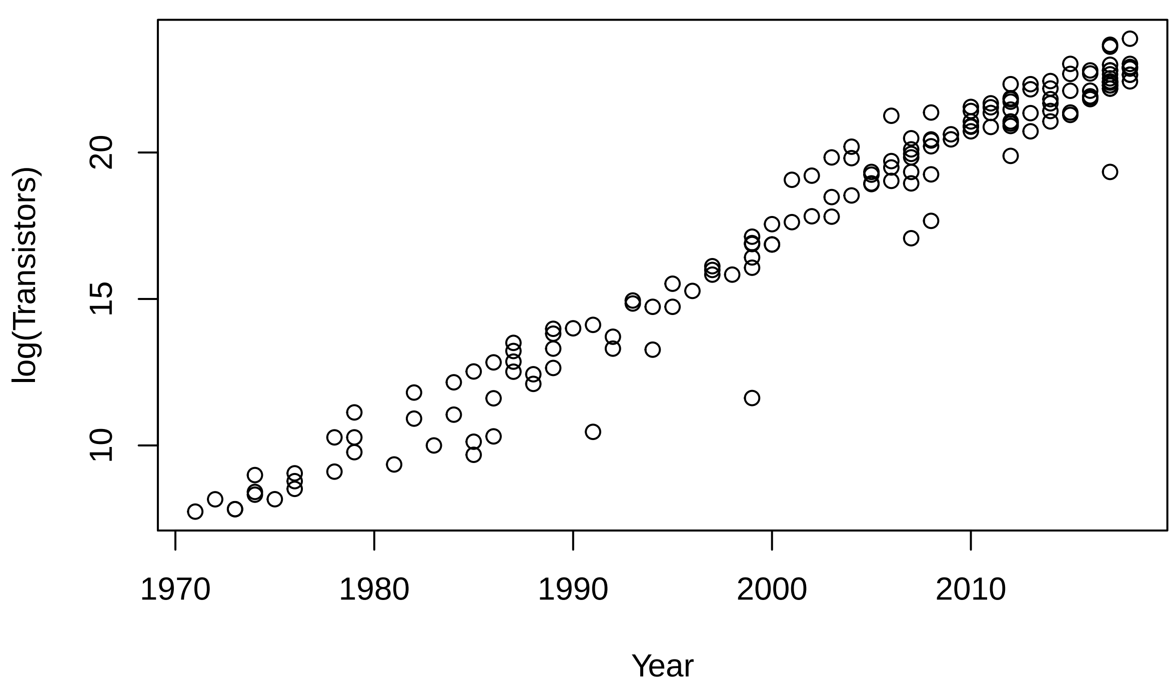
According to the formular for Moore's law and take log on T, we get:

$log(T) \approx log(C * 2^{A/2}) \rightarrow log(T) \approx log(C) + 0.3465736 * A$

This is the form of linear regression that make logT as response variable and A as explanatory variable. `log(C)` is the intercept and coefficient of A is 0.3465736. Thus we can say logT is roughly follow a simple linear regression on A.

(ii)

```
moores_df = read.csv("mooreslawdata.csv", header=TRUE)
with(moores_df, plot(log(Transistors) ~ Year))
```



Solution for Problem (b)

(i)

```
model {
  for (i in 1:length(y)) {
  y[i] ~ dnorm(beta1 + beta2*A_CENTERED[i], sigmasqinv)
  }
  beta1 ~ dnorm(0, 1e-06)
  beta2 ~ dnorm(0, 1e-06)
  sigmasqinv ~ dgamma(0.001, 0.001)
  sigmasq <- 1/sigmasqinv
}
```

```
df <- list(y = log(moores_df$Transistors),
           A_CENTERED = moores_df$Year - mean(moores_df$Year))

lm_model = lm(y ~ A_CENTERED, data=df)
summary(lm_model)
```

```
##
## Call:
## lm(formula = y ~ A_CENTERED, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0598 -0.4443  0.0891  0.6052  2.1820
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.726825   0.074044  239.41   <2e-16 ***
## A_CENTERED   0.342122   0.005436   62.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9395 on 159 degrees of freedom
## Multiple R-squared:  0.9614, Adjusted R-squared:  0.9612
## F-statistic:  3960 on 1 and 159 DF,  p-value: < 2.2e-16
```

- β estimates are about 0.3 to 18, we choose to set initial β values at ± 200
- Regression error variance σ^2 estimate is about $(0.94^2) \approx 0.9$, we choose to set initial σ^2 values of 0.01 and 100

```
gelman.diag(coef_sample, autoburnin=FALSE)
```

```
## Potential scale reduction factors:
##
##         Point est. Upper C.I.
## beta1            1          1
## beta2            1          1
## sigmasq          1          1
##
## Multivariate psrf
##
## 1
```

Gelman-Rubin statistic for $\beta1$, $\beta2$ and σ^2 are all 1, thus we can declare convergence of them.

(ii)

```r
summary(coef_sample)
```

```
##
## Iterations = 1001:3000
## Thinning interval = 1
## Number of chains = 4
## Sample size per chain = 2000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##            Mean       SD  Naive SE Time-series SE
## beta1   17.7275 0.074564 8.337e-04      8.371e-04
## beta2    0.3420 0.005416 6.055e-05      6.007e-05
## sigmasq  0.8944 0.102059 1.141e-03      1.151e-03
##
## 2. Quantiles for each variable:
##
##            2.5%     25%     50%     75%   97.5%
## beta1   17.5815 17.6782 17.7267 17.7783 17.8735
## beta2    0.3313  0.3384  0.3421  0.3457  0.3525
## sigmasq  0.7149  0.8245  0.8870  0.9575  1.1136
```

(iii)

```r
beta2_sample = as.matrix(coef_sample)[,"beta2"]
```

Mean of slope,

```r
mean(beta2_sample)
```

```
## [1] 0.3420225
```

95% posterior credible interval of slope,

```r
quantile(beta2_sample, c(0.025, 0.975))
```

```
##      2.5%     97.5%
## 0.3312595 0.3525215
```

The interval contains value (0.3465736) determined in part (a).

(iv)

```r
beta1_sample = as.matrix(coef_sample)[,"beta1"]
```

Mean of intercept,

```r
mean(beta1_sample)
```

```
## [1] 17.72748
```

95% posterior credible interval of intercept,

```r
quantile(beta1_sample, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 17.58150 17.87346
```

Solution for Problem (c)

Solution for Problem (d)