Homework 1 Solutions

Pratik Dahal

October 15, 2021

Problem 1:

Posterior mean of θ is 1890 (approximately). Moreover, consider the following figure:

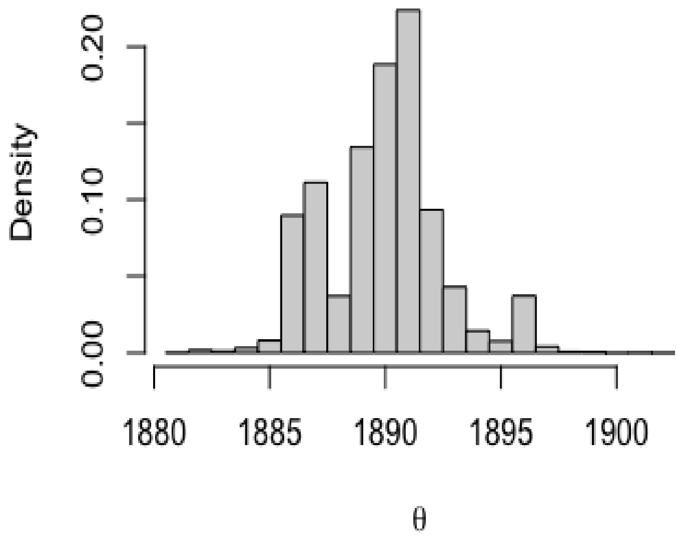


Figure 1: Histogram of the posterior distribution.

Lastly, the 90% Credible Interval is as follows:

2.5%	**97.5**%
1886	1896

□

Problem 2:

Estimate of $\lambda_1 = 3.109547$, and $\lambda_2 = 0.9237612$. (See *Appendix*) □

Problem 3:

The scatterplot of λ_2 against λ_1 for the initial SIR sample is given below. Also, the points resampled at the second stage of SIR is highlighted in red. (See *Appendix*) □

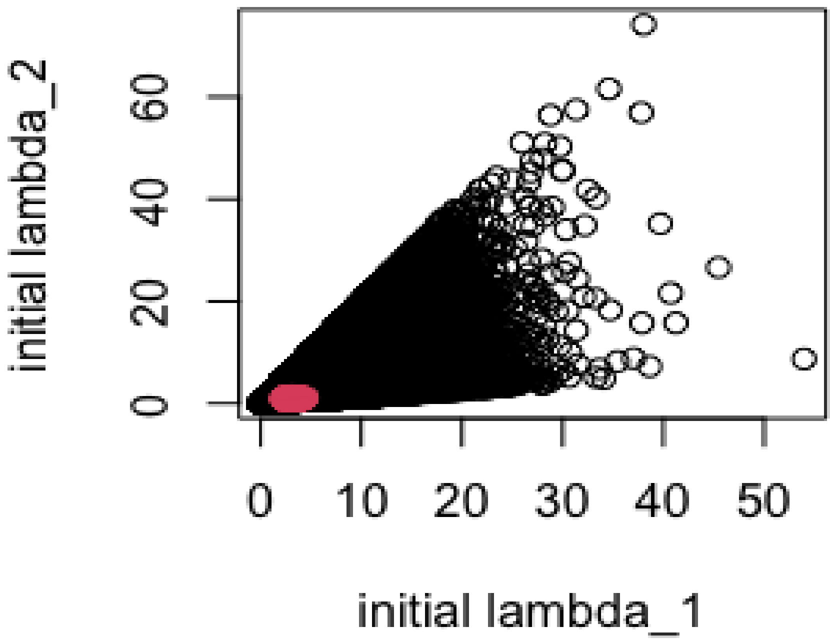


Figure 2: Scatterplot of λ_2 against λ_1.

Problem 4:

(See *Appendix*)

Initial sample size = 1000000

Resampling sample size = 10000

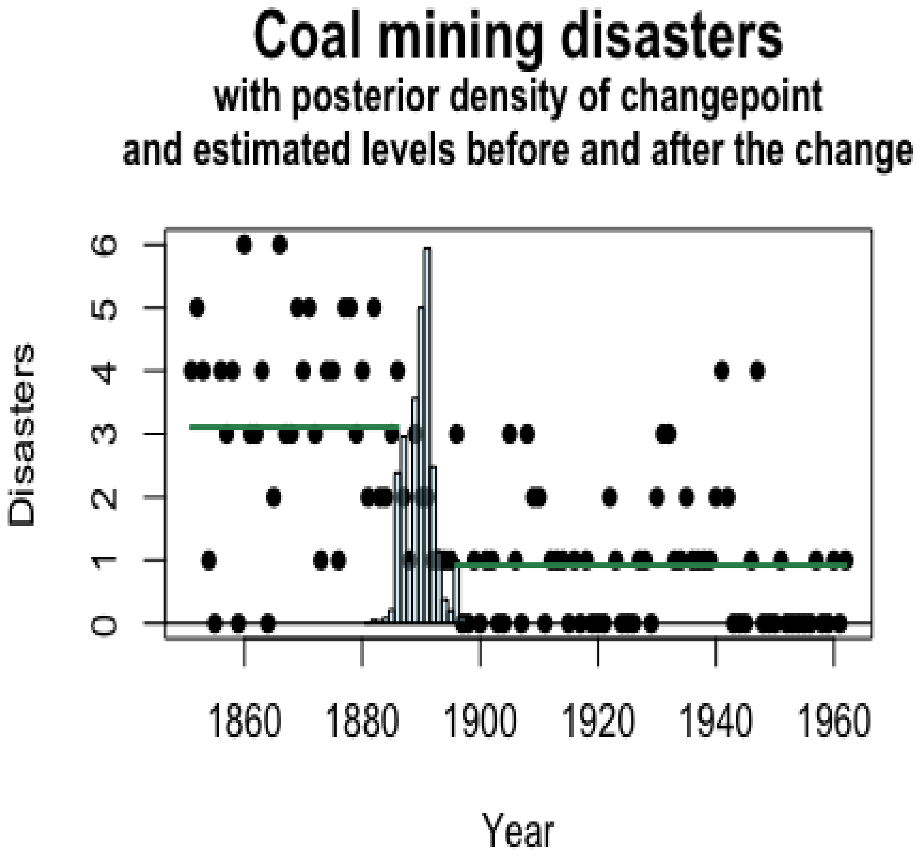
Number of unique points in the final sample = 20

Highest observed frequency in the final sample = 2239 (year = 1891)

Effective Sample Size = 1800 (approximately) □

Problem 5:

SIR model involves using a fixed number of resampling sample size to get approximate i.i.d draws from the target distribution. Computing the normalized weights gives us the power to approximate the target distribution upto a proportionality constant which, in turn, is very useful to reasonably estimate the posterior distribution upto some accuracy. Ideally, to achieve "good" convergence, the ratio between resampling sample size and the initial sample size must be small. This is true in our case since the ratio is equal to 0.01. Moreover, The effective sample size of 1800 (approximately). We want this quantity to be as high as possible as this indicates that the envelope used to sample from the target distribution is robust. The change point is similar to the one that we found in class. Lastly, the combined plot is as follows: □



Appendix

```r
######
###### Homework 1
###### Credit for the R code goes to Dr. Petris.
######

####
#### Coal mining disasters
####

### Reading the data
dat <- read.table("/Users/pratikdahal/Desktop/comp_stat/coal.dat", header = TRUE)
y <- dat$disasters
n <- length(y)

### Visualizing the data
plot(1851:1962, y, pch = 20, main = "Coal␣mining␣disasters",
     xlab = "Year", ylab = "Disasters")

# Initial sample size (Problem 4)
MC <- 1000000L
set.seed(15)

out_1 <- matrix(0, nrow = MC, ncol = 5)
out_1[, 1] <- sample(112, size = MC, replace = TRUE)  #vector theta
out_1[, 5] <- rgamma(MC, shape = 10, rate = 10)  #vector a
out_1[, 4] <- exp(runif(MC, min = log(1/8), max = log(2)))  #vector alpha
out_1[, 2] <- rgamma(MC, shape = 3, rate = out_1[, 5])   #vector lambda_1
out_1[, 3] <- out_1[, 2] * out_1[, 4]  #vector lambda_2

### Compute the un-normalized weights
llik <- apply(out_1, 1, function(u) sum(dpois(y, rep(u[2:3], c(u[1], n-u[1])), log = TRUE)))
llik <- llik - max(llik) # in order to avoid overflow when exponentiating
w_star <- exp(llik)

###  Normalize the weights
w <- w_star / sum(w_star)
1 / sum(w^2) # Effective Sample Size (Problem 4)

###  Resample, with probabilities equal to the weights (Problem 4)
MC_2 <- 10000L
out_2 <- out_1[sample(MC, size = MC_2, replace = TRUE, prob = w), ]
theta <- out_2[, 1] + 1850

### Summaries...
range(theta)
# Plotting histogram of the posterior (Problem 1)
hist(theta, breaks = 0.5 + (min(theta)-1):(max(theta)+1),
     prob = TRUE, xlab = expression(theta), main = "")
quantile(theta, c(0.025, 0.975)) # 95% Bayesian Credible Interval (Problem 1)

length(unique(theta)) # number of unique points (Problem 4)
sort(unique(theta))
table(theta)
mean(theta) # estimating posterior mean (Problem 1)
max(table(theta))  # maximum frequency  (Problem 4)


par(mar = c(5, 4, 5, 2) + 0.1)
plot(1851:1962, y, pch = 16, main = "", xlab = "Year", ylab = "Disasters")
h <- hist(theta, breaks = 0.5 + (min(theta)-1):(max(theta)+1), plot = FALSE)
abline(h = 0)
usr <- par("usr")
col2rgb("lightblue")
myblue <- rgb(173, 216, 230, alpha = 110, max = 255)
rect(h$mids-0.5, 0, h$mids+0.5, usr[4] * h$density / (1.05 * max(h$dens)),col = myblue)
```

```r
mtext("Coal␣mining␣disasters", line = 3, font = 2, cex = 1.5)
mtext("with␣posterior␣density␣of␣changepoint", line = 2, font = 2)
mtext("and␣estimated␣levels␣before␣and␣after␣the␣change", line = 1, font = 2)

# Estimated values from the posterior (Problem 2)
lambda_1 <- mean(out_2[, 2])
lambda_2 <- mean(out_2[, 3])
segments(1851, lambda_1, 1886, lambda_1, lwd = 2, col = "seagreen")
segments(1896, lambda_2, 1962, lambda_2, lwd = 2, col = "seagreen")

# Plot for (Problem 3).
plot(out_1[, 2],out_1[, 3], main = "", xlab = "initial␣lambda_1", ylab = "initial␣lambda_2")
points(out_2[, 2],out_2[, 3],col=2)
```