

# Gibbs Sampling

## Case Study: The Changepoint Problem

Math 392

## For later use:

The random variable  $X$  is *Poisson* if,

$$f(x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}; \quad \lambda > 0, x \in 0, 1, 2, \dots$$

The random variable  $X$  is *Gamma* if,

$$f(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; \quad \alpha > 0, \beta > 0, x \in 0, \infty$$

where  $E(X) = \frac{\alpha}{\beta}$  and  $Var(X) = \frac{\alpha}{\beta^2}$ .

# Markov Chain Monte Carlo

A collection of algorithms that allow you to take random (Monte Carlo) samples from distributions using a Markov chain.

# Gibbs Sampling

Suppose you wish to sample from a joint pdf with density  $f(x, y)$ , but are unable to do it directly (there is no `r_---()` function). If we can generate from the respective *conditional* distributions, we can set up a Markov chain.

Start with a reasonable  $Y_0$ , then:

1. Generate  $X_{i+1} \sim f_x(X|Y_i)$
2. Generate  $Y_{i+1} \sim f_y(Y|X_{i+1})$

Iterate these steps, and if conditions are met, the stationary distribution will be  $f(x, y)$ .

Practical details:

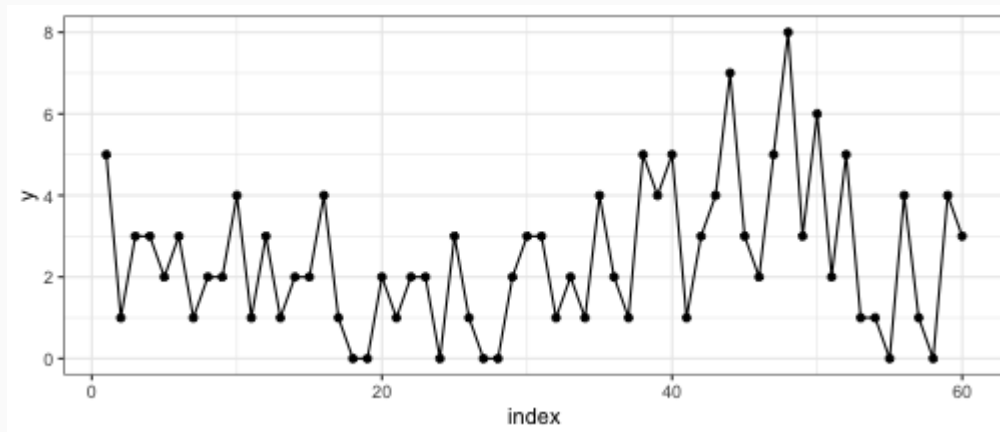
- Use a burn-in period to escape  $Y_0$ .
- Take every tenth sample to diminish Markov dependence.

# The Changepoint Problem

Consider a series of Poisson random variables,  $Y_1, Y_2, \dots, Y_n$ . At some unknown point,  $m$ , the rate of the Poisson distribution changed. That is,

$$Y_1, \dots, Y_m \sim \text{Poi}(\mu)$$
$$Y_{m+1}, \dots, Y_n \sim \text{Poi}(\lambda)$$

Say:  $\mu = 2$ ,  $\lambda = 4$  and  $m = 38$  (at  $n = 60$ ).



**Goal:** Estimate  $m$  using data and prior information.

# Plan of Attack

I. Specify prior

II. Write down full joint posterior

III. Write down joint posterior of parameters

IV. Write down conditional posteriors

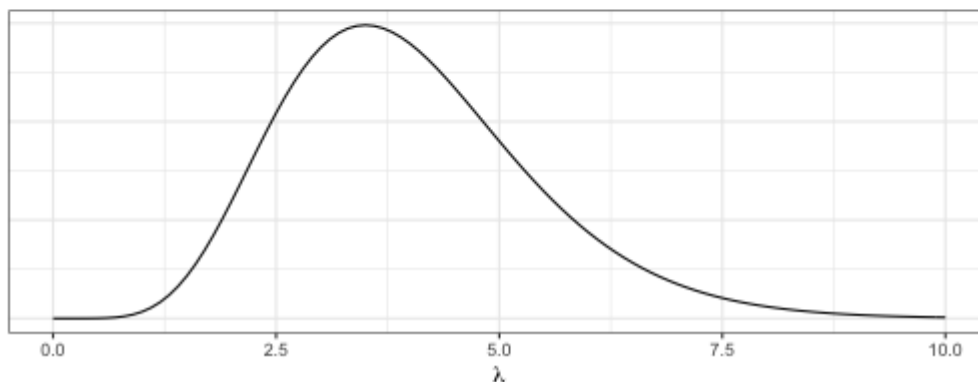
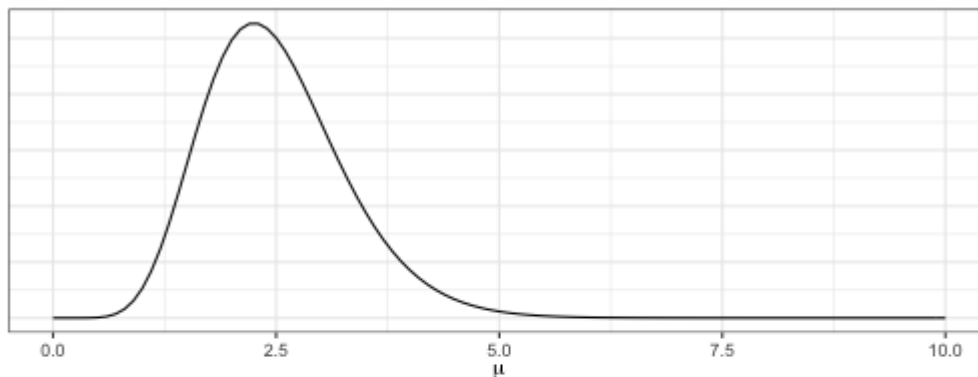
V. Run Gibbs Sampler to draw from joint posterior then assess convergence and Markov dependence

VI. Formulate conclusions

# I. Specifying Priors

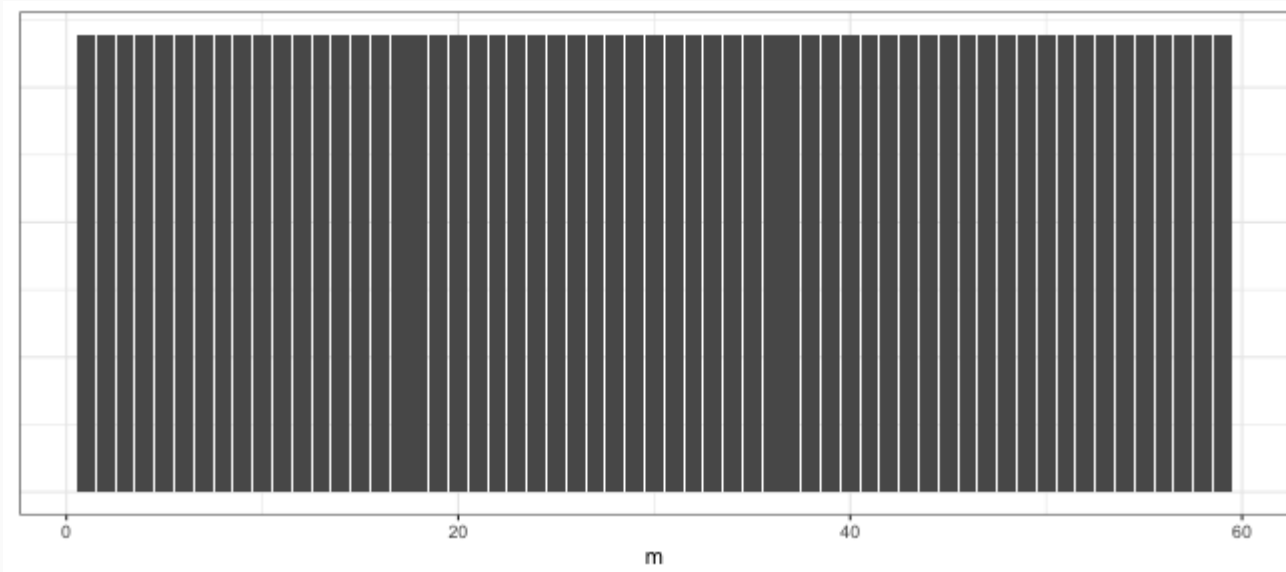
$$\mu \sim \text{Gamma}(\alpha = 10, \beta = 4)$$

$$\lambda \sim \text{Gamma}(\nu = 8, \phi = 2).$$



# I. Specifying Priors, cont.

$$m \sim \text{Unif}\{1, 2, \dots, n - 1\}$$





## II. Full Joint Distribution

If they were all independent:

$$f(Y_\mu, Y_\lambda, \mu, \lambda, m) = f(Y_\mu)f(Y_\lambda)f(\mu)f(\lambda)f(m)$$

Correct joint distribution:

$$\begin{aligned} &= f(Y_\mu | \mu, m) f(Y_\lambda | \lambda, m) f(\mu) f(\lambda) f(m) \\ &= \prod_{i=1}^m \left( \frac{\mu^{Y_i} e^{-\mu}}{Y_i!} \right) \prod_{j=m+1}^n \left( \frac{\lambda^{Y_j} e^{-\lambda}}{Y_j!} \right) \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta\mu} \frac{\phi^\nu}{\Gamma(\nu)} \lambda^{\nu-1} e^{-\phi\lambda} \frac{1}{n-1} \end{aligned}$$

### III. Joint Posterior of Parameters

From the definition of conditional probability:

$$f(\mu, \lambda, m | Y_\mu, Y_\lambda) = \frac{f(Y_\mu, Y_\lambda, \mu, \lambda, m)}{f(Y_\mu, Y_\lambda)}$$

We can then pull out the terms,  $c$ , that are not a function of the parameters.

$$\begin{aligned} &= c f(Y_\mu, Y_\lambda, \mu, \lambda, m) \\ &= c \prod_{i=1}^m \left( \frac{\mu^{Y_i} e^{-\mu}}{Y_i!} \right) \prod_{j=m+1}^n \left( \frac{\lambda^{Y_j} e^{-\lambda}}{Y_j!} \right) \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta\mu} \frac{\phi^\nu}{\Gamma(\nu)} \lambda^{\nu-1} e^{-\phi\lambda} \frac{1}{n-1} \\ &= c \prod_{i=1}^m (\mu^{Y_i} e^{-\mu}) \prod_{j=m+1}^n (\lambda^{Y_j} e^{-\lambda}) \mu^{\alpha-1} e^{-\beta\mu} \lambda^{\nu-1} e^{-\phi\lambda} \\ &= c \mu^{\sum_{i=1}^m Y_i} e^{-m\mu} \lambda^{\sum_{j=m+1}^n Y_j} e^{-(n-m)\lambda} \mu^{\alpha-1} e^{-\beta\mu} \lambda^{\nu-1} e^{-\phi\lambda} \\ &= c \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta+m)\mu} \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi+n-m)\lambda} \end{aligned}$$

## IV. Conditional Posterior for $m$

We again use the definition of conditional probability:

$$\begin{aligned} f(m \mid \mu, \lambda, Y_\mu, Y_\lambda) &= \frac{f(\mu, \lambda, m \mid Y_\mu, Y_\lambda)}{f(\mu, \lambda \mid Y_\mu, Y_\lambda)} \\ &= \frac{c \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda}}{\sum_{k=1}^{n-1} c \mu^{\alpha + \sum_{i=1}^k Y_i - 1} e^{-(\beta + k)\mu} \lambda^{\nu + \sum_{j=k+1}^n Y_j - 1} e^{-(\phi + n - k)\lambda}} \end{aligned}$$

Once we cancel out the  $c$ , this serves as our discrete posterior distribution on  $m$ .

## IV. Conditional Posterior for $\mu$

We start by finding the normalizing constant that we'll use in the denominator,  $f(\lambda, m \mid Y_\mu, Y_\lambda)$ .

$$\begin{aligned} &= \int_0^\infty f(\mu, \lambda, m \mid Y_\mu, Y_\lambda) \, d\mu \\ &= \int_0^\infty c \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda} \, d\mu \\ &= c \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda} \int_0^\infty \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \, d\mu \\ &= c \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda} \frac{\Gamma(\alpha + \sum_{i=1}^m Y_i)}{(\beta + m)^{\alpha + \sum_{i=1}^m Y_i}} \int_0^\infty \frac{(\beta + m)^{\alpha + \sum_{i=1}^m Y_i}}{\Gamma(\alpha + \sum_{i=1}^m Y_i)} \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \, d\mu \\ &= c \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda} \frac{\Gamma(\alpha + \sum_{i=1}^m Y_i)}{(\beta + m)^{\alpha + \sum_{i=1}^m Y_i}} \end{aligned}$$

## IV. Conditional Posterior for $\mu$ , cont.

With this normalizing constant in hand, we can write out the form of the posterior for  $\mu$ ,  $f(\mu \mid \lambda, m, Y_\mu, Y_\lambda)$ :

$$\begin{aligned} &= \frac{f(\mu, \lambda, m, \mid Y_\mu, Y_\lambda)}{f(\lambda, m \mid Y_\mu, Y_\lambda)} \\ &= \frac{c \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda}}{c \lambda^{\nu + \sum_{j=m+1}^n Y_j - 1} e^{-(\phi + n - m)\lambda} \frac{\Gamma(\alpha + \sum_{i=1}^m Y_i)}{(\beta + m)^{\alpha + \sum_{i=1}^m Y_i}}} \\ &= \frac{(\beta + m)^{\alpha + \sum_{i=1}^m Y_i}}{\Gamma(\alpha + \sum_{i=1}^m Y_i)} \mu^{\alpha + \sum_{i=1}^m Y_i - 1} e^{-(\beta + m)\mu} \end{aligned}$$

Which is recognizable as the pdf of a Gamma random variable. Therefore,

$$\mu \mid m, Y_\mu \sim \text{Gamma}\left(\alpha + \sum_{i=1}^m Y_i, \beta + m\right)$$

## IV. Posterior for $\lambda$

Same story as for  $\mu$ .

$$\lambda \mid m, Y_\lambda \sim \text{Gamma}(\nu + \sum_{i=m+1}^n Y_i, \phi + n - m)$$

## V. Gibbs Sampler

Form a Markov chain that begins by initializing a value for  $m_{j-1}$ , then iterates through the following three steps many times.

1. Sample  $\mu_j$  from  $\text{Gamma}(\alpha + \sum_{i=1}^{m_{j-1}} Y_i, \beta + m_{i-j})$
2. Sample  $\lambda_j$  from  $\text{Gamma}(\nu + \sum_{i=m_{j-1}+1}^n Y_i, \phi + n - m_{j-1})$
3. Sample  $m_j$  from  $f(m \mid \mu_j, \lambda_j, Y_{\mu_j}, Y_{\lambda_j})$

Our specific scenario:  $n = 60$ ,  $\alpha = 10$ ,  $\beta = 4$ ,  $\nu = 8$ , and  $\phi = 2$ .

```

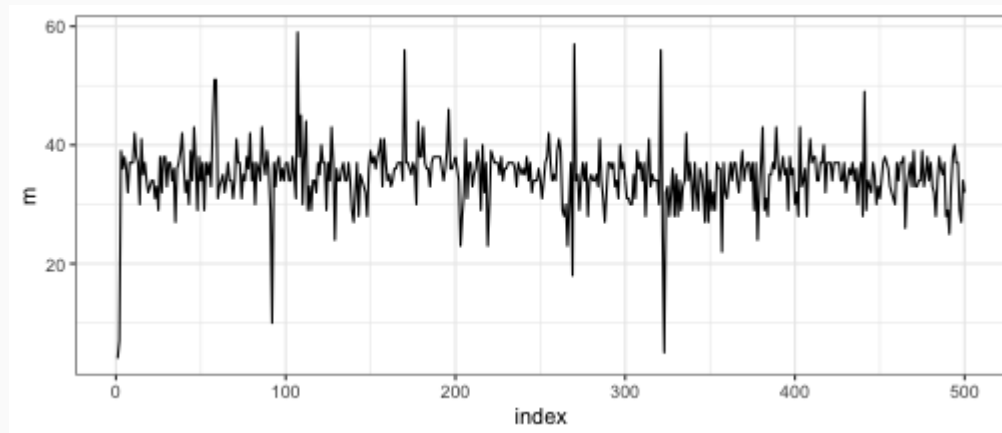
it <- 50000
post_samples <- matrix(rep(NA, it * 3), ncol = 3)
colnames(post_samples) <- c("mu", "lambda", "m")
m_j <- 2 # initialize m
for (j in 1:it) {
  # sample mu
  mu_j <- rgamma(1, alpha + sum(y[1:m_j]), beta + m_j)
  # sample lambda
  lambda_j <- rgamma(1, nu + sum(y[(m_j+1):n]), phi + (n -
  # sample m
  m_vec <- rep(NA, n - 1)
  for (k in 1:(n - 1)) {
    m_vec[k] <- mu_j^(alpha + sum(y[1:k]) - 1) *
      exp(-(beta + k) * mu_j) *
      lambda_j^(nu + sum(y[(k+1):n]) - 1) *
      exp(-(phi + n - k) * lambda_j)
  }
  p <- m_vec/sum(m_vec)
  m_j <- sample(1:(n - 1), size = 1, prob = p)
  # store results
  post_samples[j, "mu"] <- mu_j
  post_samples[j, "lambda"] <- lambda_j
  post_samples[j, "m"] <- m_j
}

```



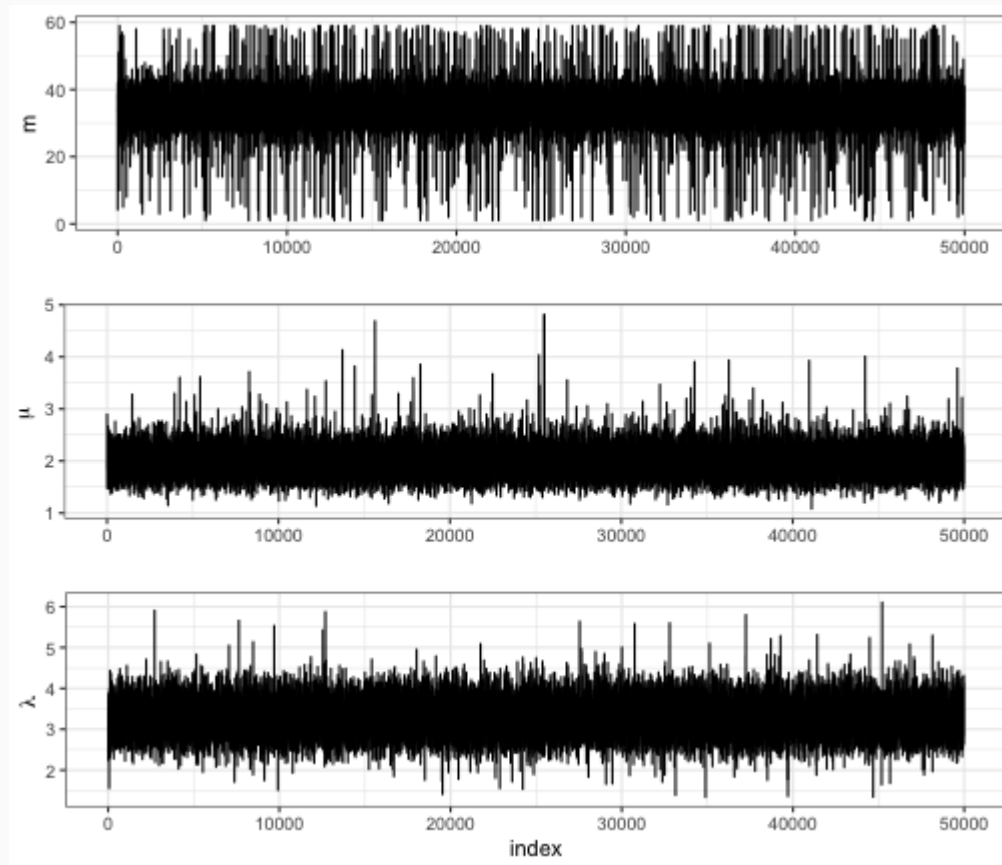
# V. Convergence

The first 500  $m_j$ :



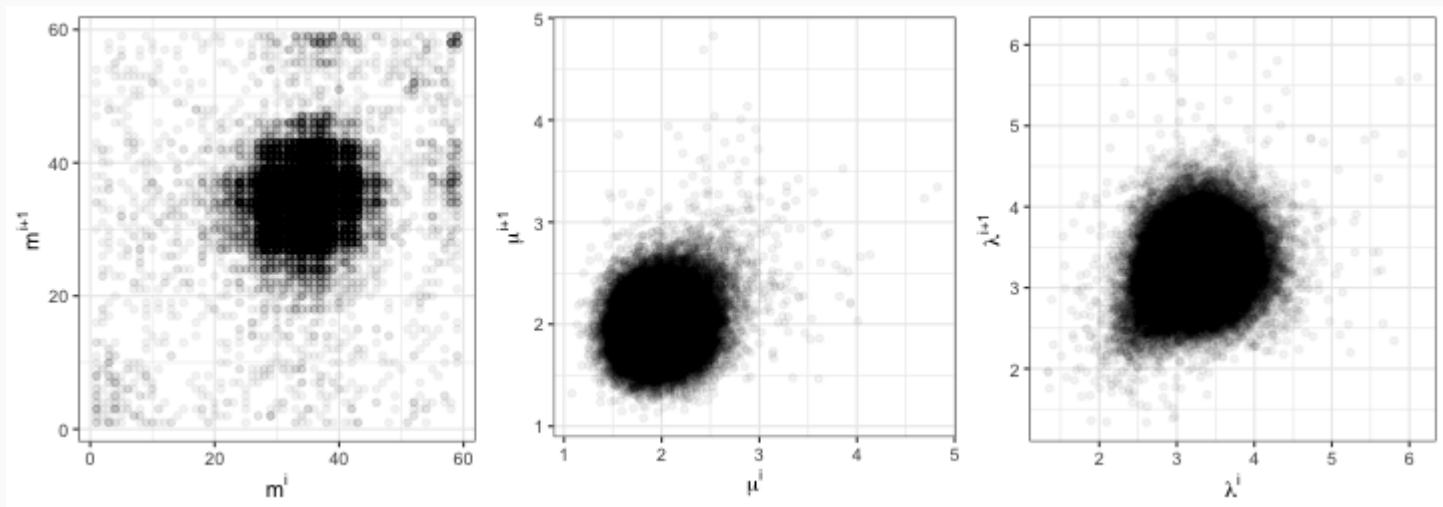
## V. Convergence, cont.

All 50,000:



## V. Thinning

To investigate serial correlation, we plot each sampled parameter against its subsequent draw.

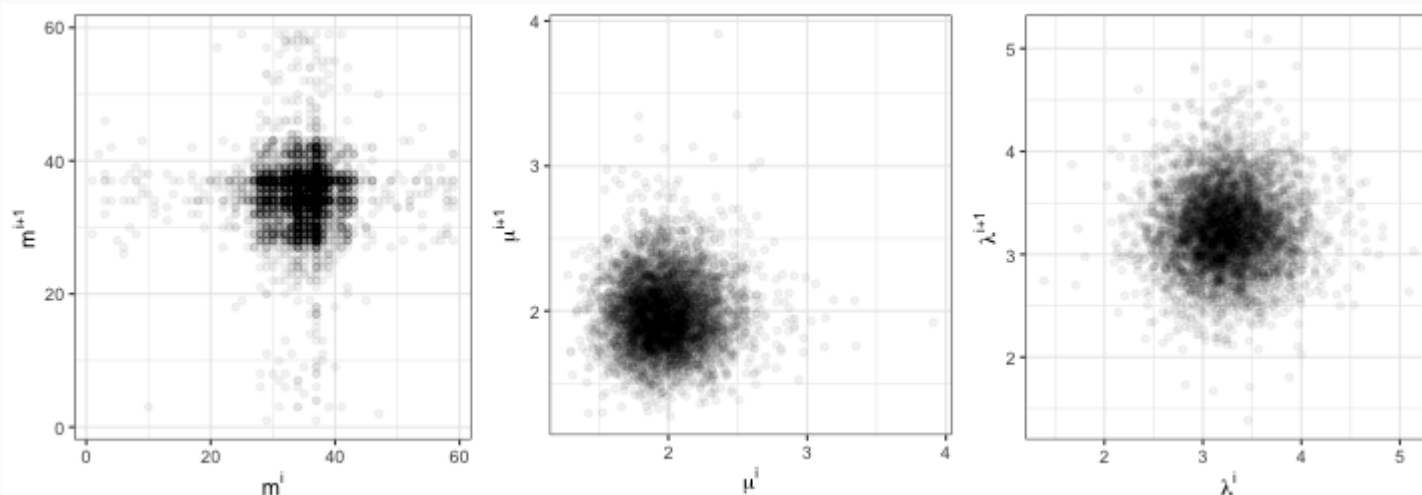


Correlations: 0.196, 0.157, and 0.158.

## V. Thinning

We can remove the Markov dependence by retaining only every tenth sample.

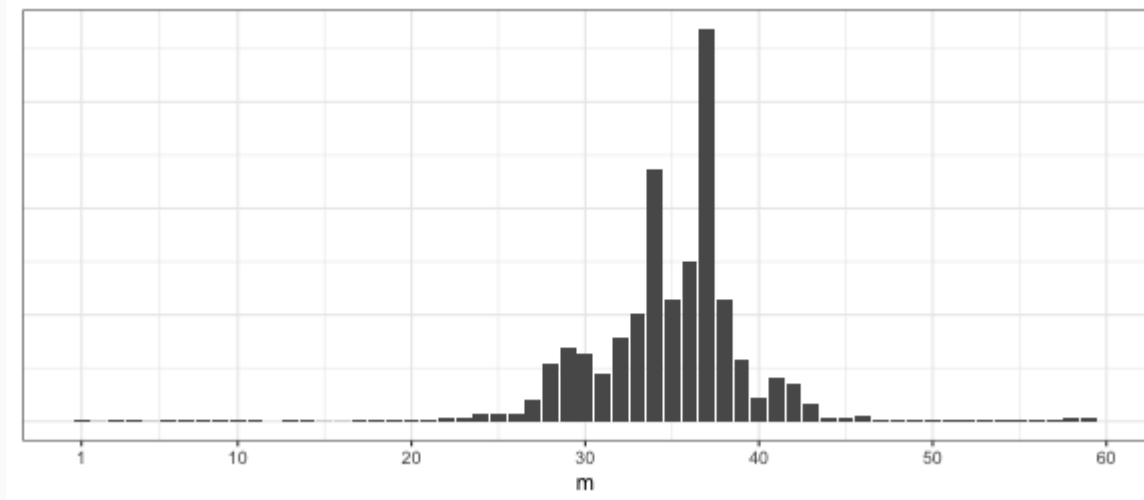
```
df <- df %>%  
  slice(seq(from = 1, to = nrow(df), by = 10))
```



Correlation: 0.002, 0.005, and 0.014.

## VI. Conclusions

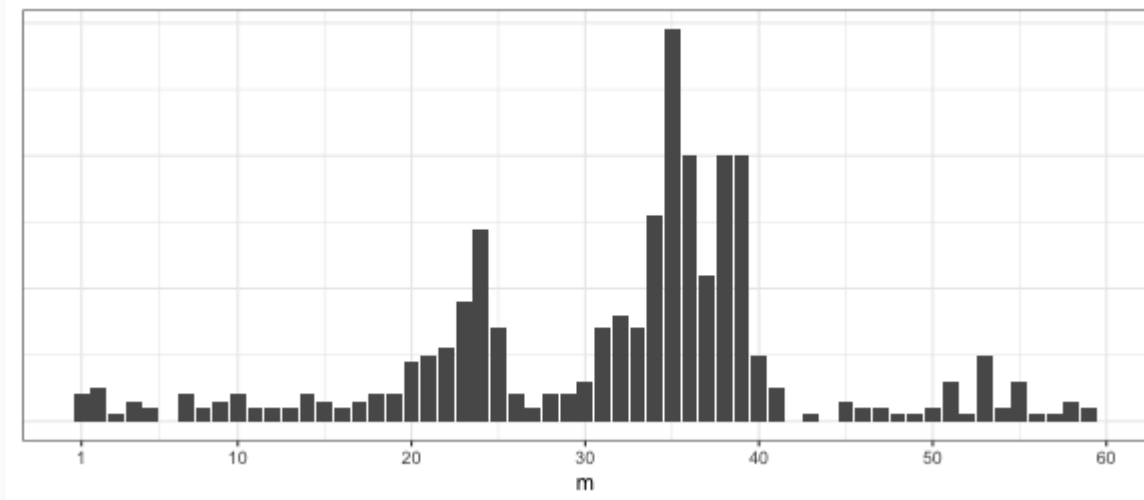
What is our updated best guess for when the changepoint occurred?  
Consider the (thinned) posterior distribution on  $m$ .



Those two prominent modes are at 34 and 37, so those are probably fine point estimates for  $m$  based on this data and your prior information.

## VI. Conclusions, cont.

Consider our this final posterior changes if we use a different seed / get a different data set  $Y$ .



With a flat prior on  $m$ , the posterior will be drawing most of its structure from the data, which, at these sample sizes, are still subject to considerable sampling variability.