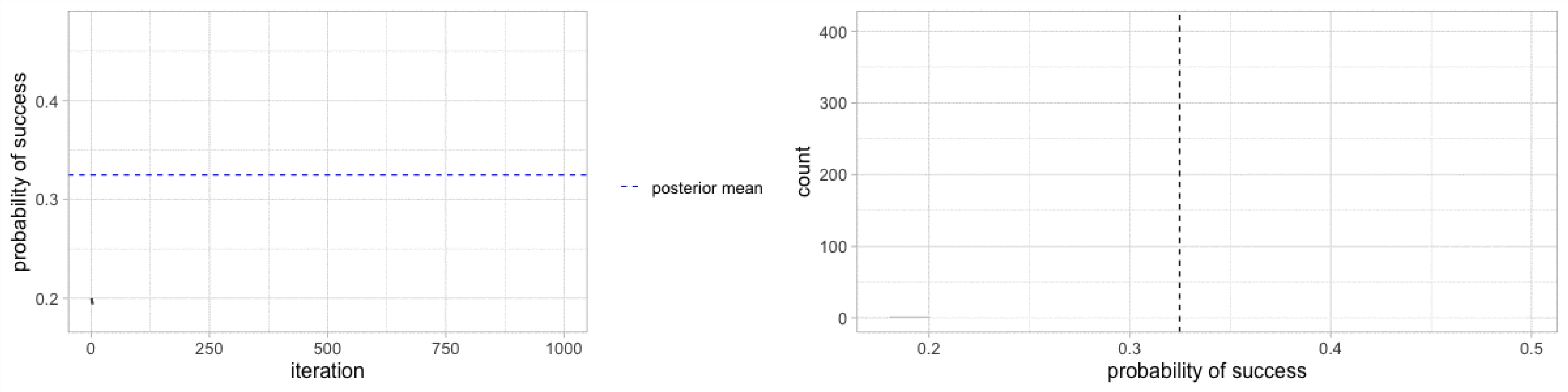
The Gibbs sampler

Stat 340: Bayesian Statistics

1. Gibbs sampler
2. Convergence checks
3. Inference using MCMC draws

(Problem topics 8, 10, 11)

Metropolis algorithm



Your turn

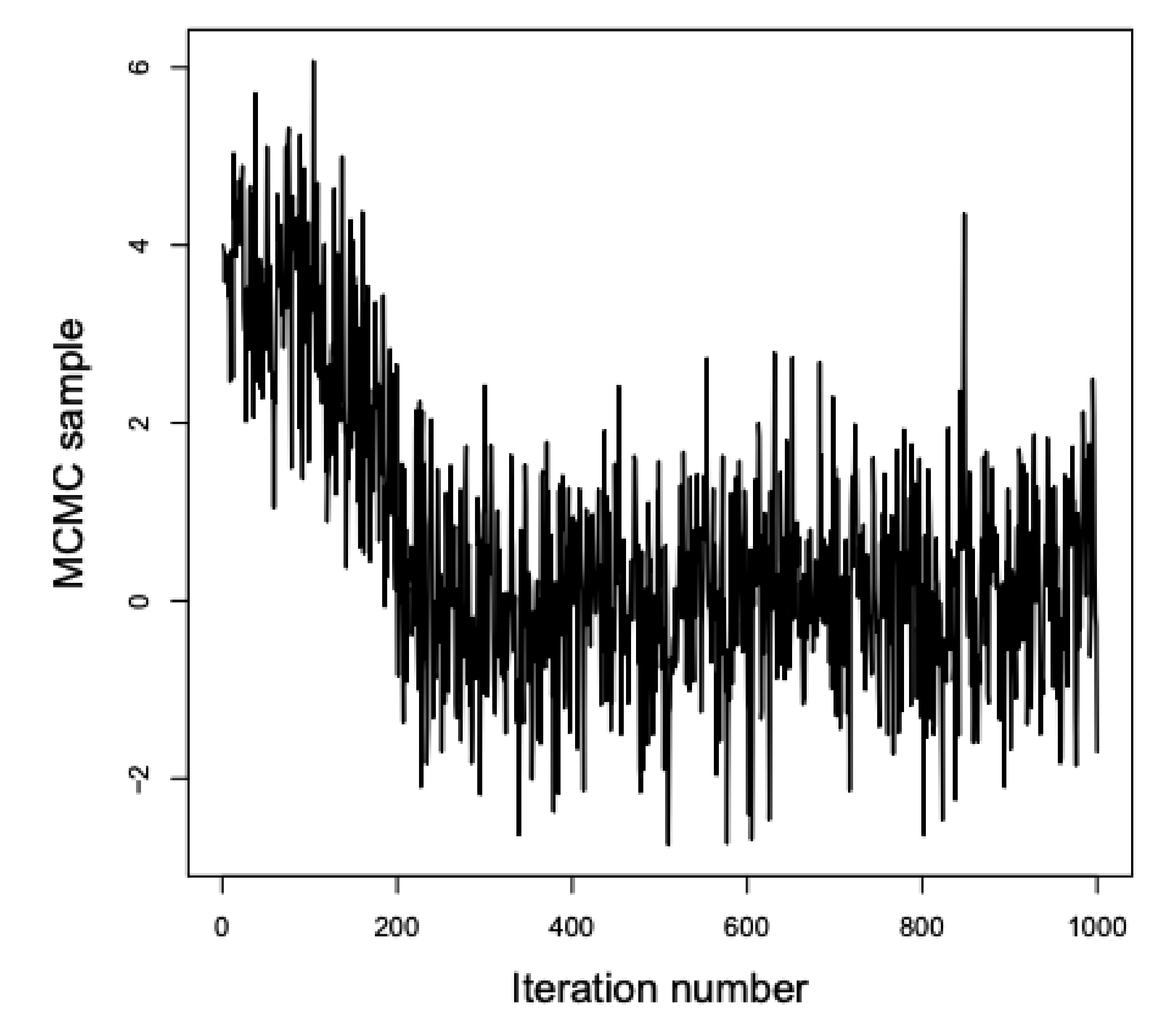
Take a look at the four trace plots provided as an example. For each, determine

1. whether the chain converged
2. roughly how many iterations it took to converge

03:00

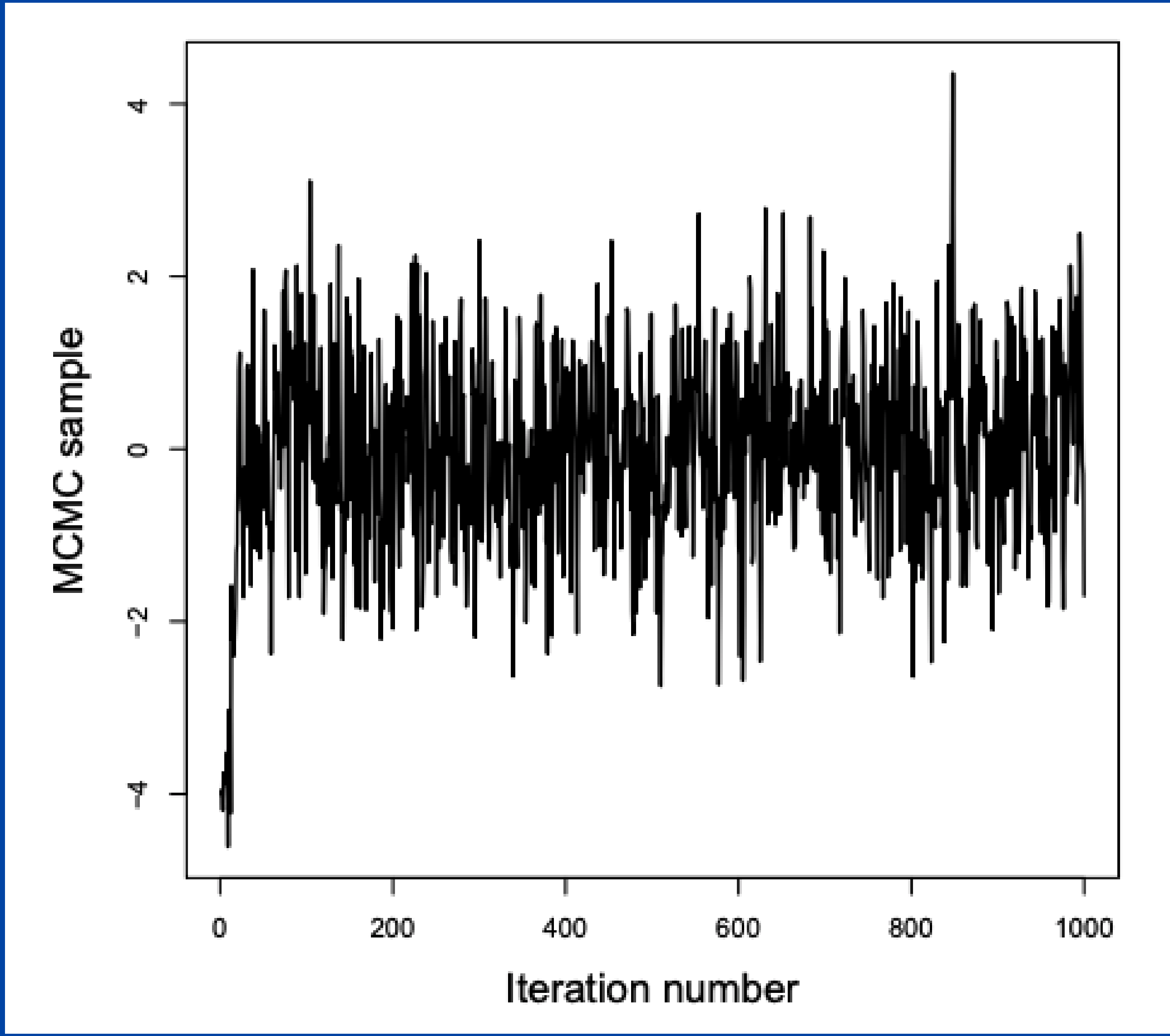
1. Did the chain converge?

2. If so, how many iterations did it take?



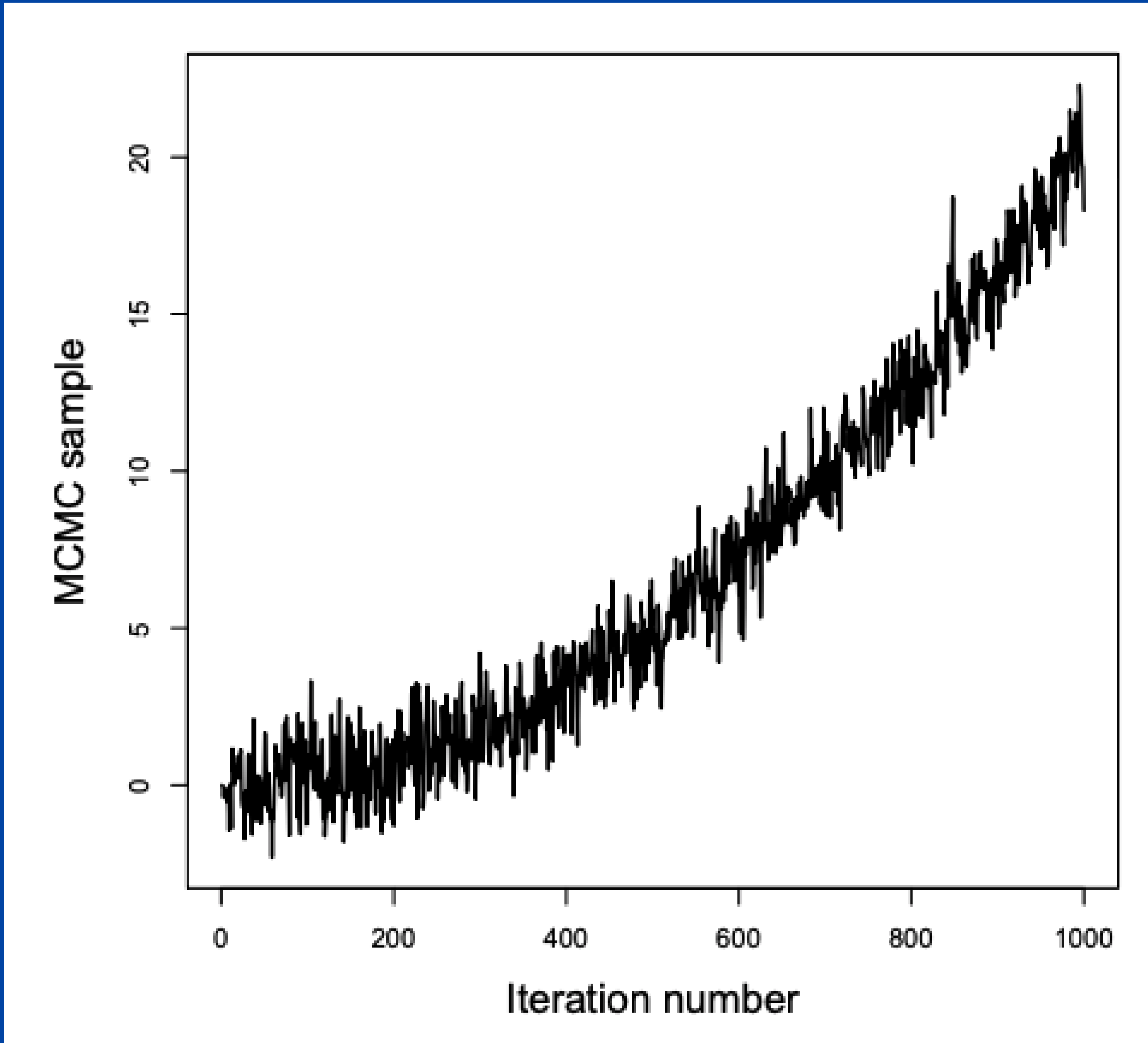
1. Did the chain converge?

2. If so, how many iterations did it take?



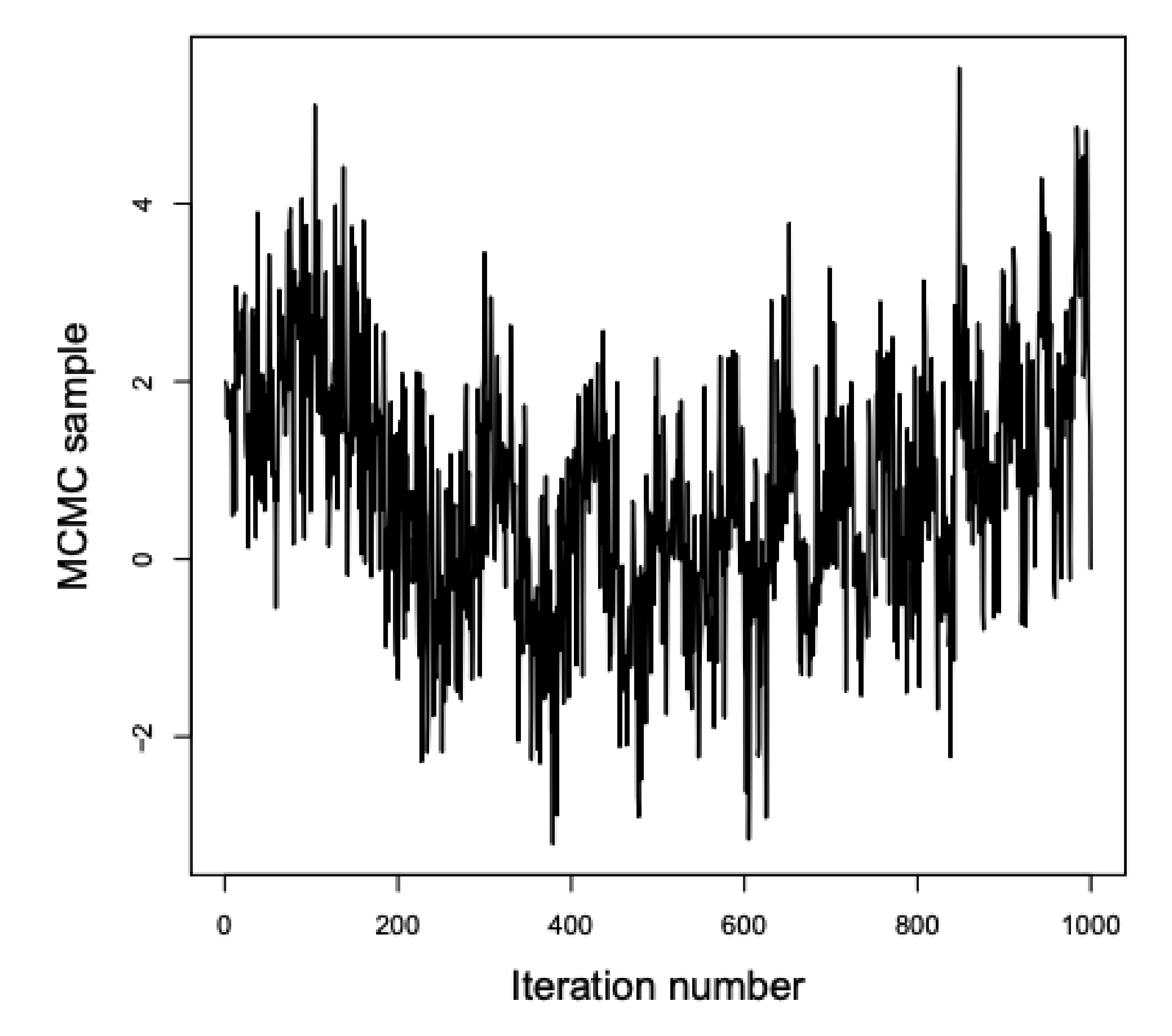
1. Did the chain converge?

2. If so, how many iterations did it take?



1. Did the chain converge?

2. If so, how many iterations did it take?



Simulating from a discrete bivariate distribution

Example: Traffic

The joint PMF of the number of cars (X) and the number of buses (Y) per signal cycle at a proposed left-turn lane is given below

X=0	Y=0	1	2
0	0.025	0.015	0.01
1	0.050	0.030	0.02
2	0.125	0.075	0.05
3	0.150	0.090	0.06
4	0.100	0.060	0.04
5	0.050	0.030	0.02

Simulation process

Condition on $X = 1$, simulate a Y

X=0	Y=0	1	2
0	0.025	0.015	0.01
1	0.050	0.030	0.02
2	0.125	0.075	0.05
3	0.150	0.090	0.06
4	0.100	0.060	0.04
5	0.050	0.030	0.02

```
sample(0:5, size = 1, prob = bivariate[[2]])
```

```
## [1] 3
```

Condition on $Y = 3$, simulate an X

X=0	Y=0	1	2
0	0.025	0.015	0.01
1	0.050	0.030	0.02
2	0.125	0.075	0.05
3	0.150	0.090	0.06
4	0.100	0.060	0.04
5	0.050	0.030	0.02

```
sample(0:2, size = 1, prob = bivariate[4,2:4])
```

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

Simulating from a bivariate continuous distribution

Change in the democratic vote

- How did democratic share of the two party vote change from 2016 to 2020?
- MIT Election Data and Science Lab has county-level election results
- We'll look at the percent change in the two-party vote

$$Y_i = 100\,(A_i/B_i - 1)$$

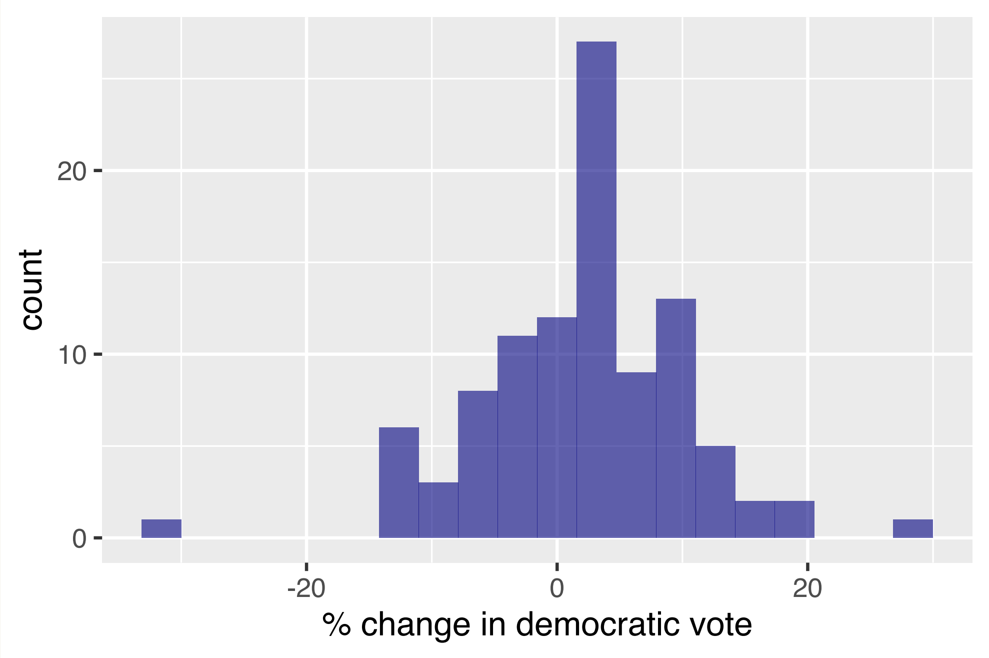
$A_i =$ % of two-party vote cast for democrats in 2020

$B_i =$ % of two-party vote cast for democrats in 2016

MIT Election Data and Science Lab, "County Presidential Election Returns 2000-2020", https://doi.org/10.7910/DVN/VOQCHQ, Harvard Dataverse, V9

Hypothetical sample

We'll work with a hypothetical sample of 100 counties



$\bar{y} = 1.96$

$n = 100$

Model

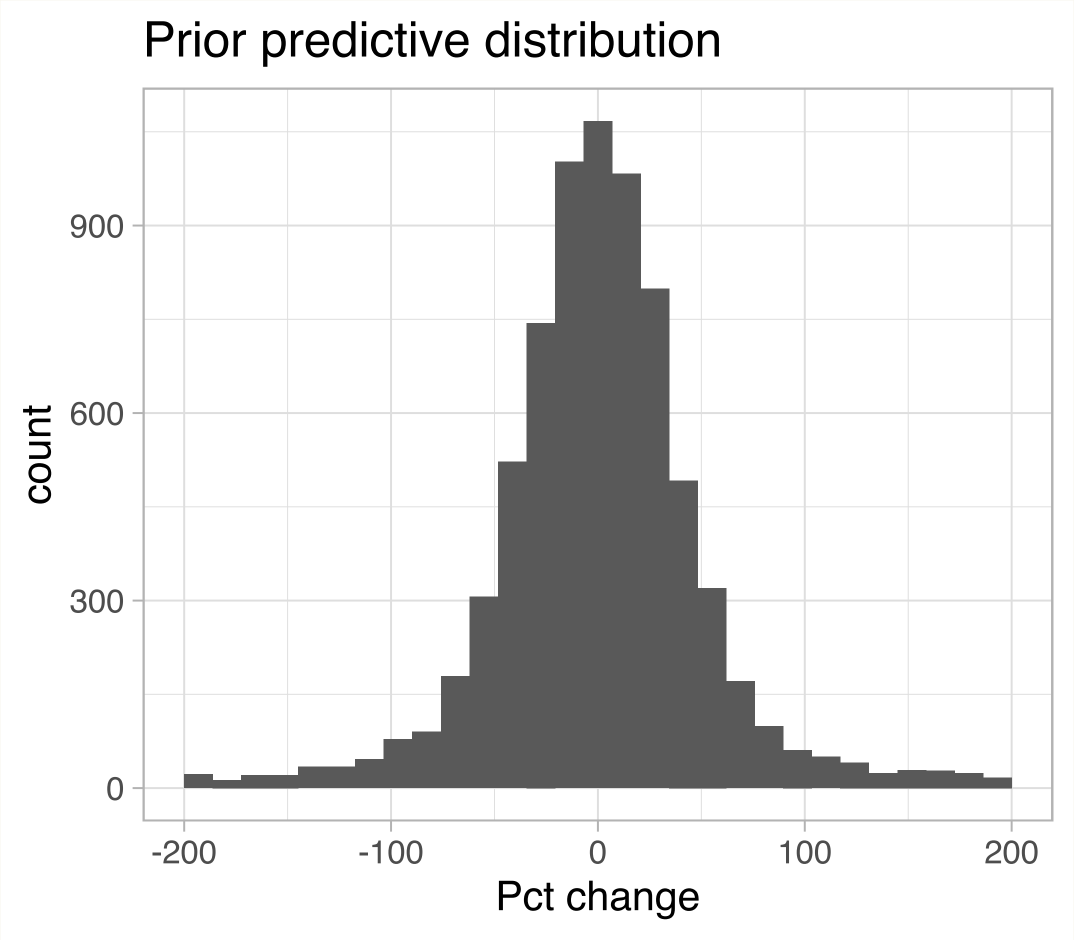
$Y_i = 100(A_i/B_i - 1)$, the percent change in support

$$\begin{aligned} Y_i|\mu,\sigma &\overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma) \\ \mu &\sim \mathcal{N}(\mu_0, \sqrt{1/\phi_0}) \\ \phi = 1/\sigma^2 &\sim \text{Gamma}(a,b) \\ \mu &\perp \phi \end{aligned}$$

$$\begin{aligned} \mu_0 &= 0 \\ \phi_0 &= 1/1000 \\ a &= 0.1 \\ b &= 0.1 \end{aligned}$$

These are very weak priors

Prior predictive check



If you don't think the priors induce a reasonable distribution on Y, then tweak the parameters (e.g. inflate σ_0)

Posterior

$$\pi(\mu, \phi | y_1, \ldots, y_n) \propto \pi(\mu)\pi(\phi) \cdot \prod_{i=1}^{n} f(y_i | \mu, \sigma^2)$$
$$\propto \exp\left[-\frac{\phi_0}{2}(\mu - \mu_0)^2\right] \cdot \phi^{a-1} \exp[-b\phi] \cdot$$
$$\prod_{i=1}^{n} \phi^{1/2} \exp\left[-\frac{\phi}{2}(y_i - \mu)^2\right]$$

How can we do this *efficiently* sample from this 2D posterior?

Two-stage Gibbs sampler

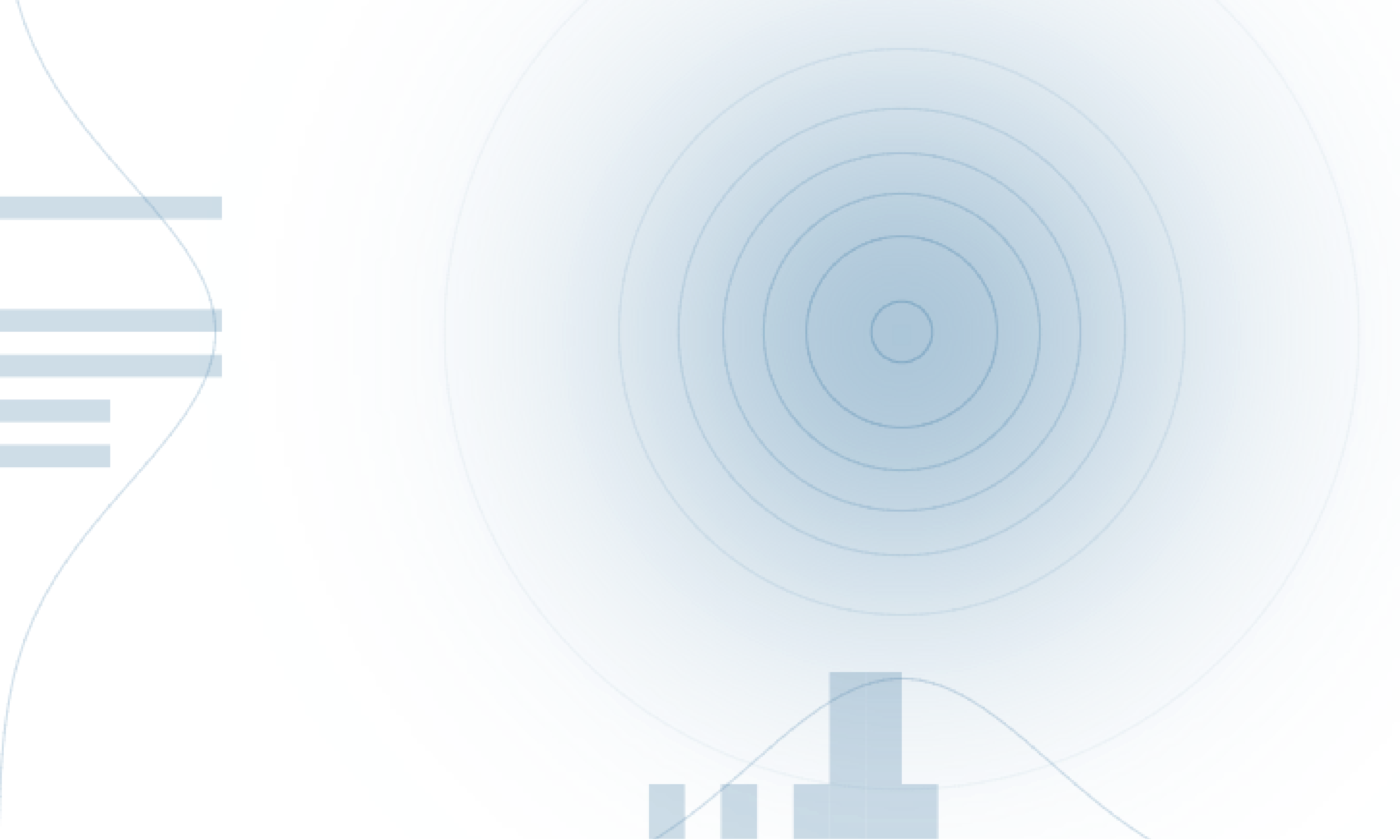
Target: samples from $\pi(\theta_1, \theta_2 | y_1, \ldots, y_n)$

Algorithm:

1. Set initial values for parameter values, $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}\right)$
2. Draw $\theta_1^{(1)}$ from $\pi(\theta_1 | \theta_2, y_1, \ldots, y_n)$
3. Draw $\theta_2^{(1)}$ from $\pi(\theta_2 | \theta_1, y_1, \ldots, y_n)$
4. Repeat steps 2-3 S times

After convergence, draws $\left(\theta_1^{(k)}, \theta_2^{(k)}\right)$ are from the posterior distribution

Animated Gibbs sampler



Your turn

In our example, $\boldsymbol{\theta} = (\mu, \sigma^2)$

Discuss with your neighbor **how** you would find the following conditional posterior distributions from the joint posterior:

1. $\pi(\mu|\phi, y_1, \ldots, y_n)$

2. $\pi(\phi|\mu, y_1, \ldots, y_n)$

02:00

Full conditional distributions

$$
\begin{aligned}
\pi(\phi|y_1,\ldots,y_n,\mu) &\propto \pi(\phi) f(y_1,\ldots,y_n|\phi,\mu) \\
&\propto \phi^{a-1}\exp[-b\phi]\cdot\prod_{i=1}^{n}\phi^{1/2}\exp\left[-\frac{\phi}{2}(y_i-\mu)^2\right] \\
&= \phi^{a-1}\exp[-b\phi]\cdot\phi^{n/2}\exp\left[-\frac{\phi}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right] \\
&= \phi^{(n/2+a)-1}\exp\left[-\phi\left\{\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2+b\right\}\right]
\end{aligned}
$$

Is this a distribution we have seen before?

Full conditional distributions

$$
\begin{aligned}
\pi(\mu|\phi, y_1, \ldots, y_n) &\propto \pi(\mu) \prod_{i=1}^{n} f(y_i|\mu, \sigma^2) \\
&\propto \exp\left[-\frac{\phi_0}{2}(\mu-\mu_0)^2\right] \cdot \prod_{i=1}^{n} \phi^{1/2} \exp\left[-\frac{\phi}{2}(y_i-\mu)^2\right] \\
&= \exp\left[-\frac{\phi_0}{2}(\mu-\mu_0)^2\right] \cdot \phi^{n/2} \exp\left[-\phi\left\{\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}\right] \\
&\propto \exp\left[-\frac{\phi_0+n\phi}{2}\left\{\mu - \frac{\mu_0\phi_0+n\bar{y}\phi}{\phi_0+n\phi}\right\}^2\right]
\end{aligned}
$$

Is this a distribution we know?

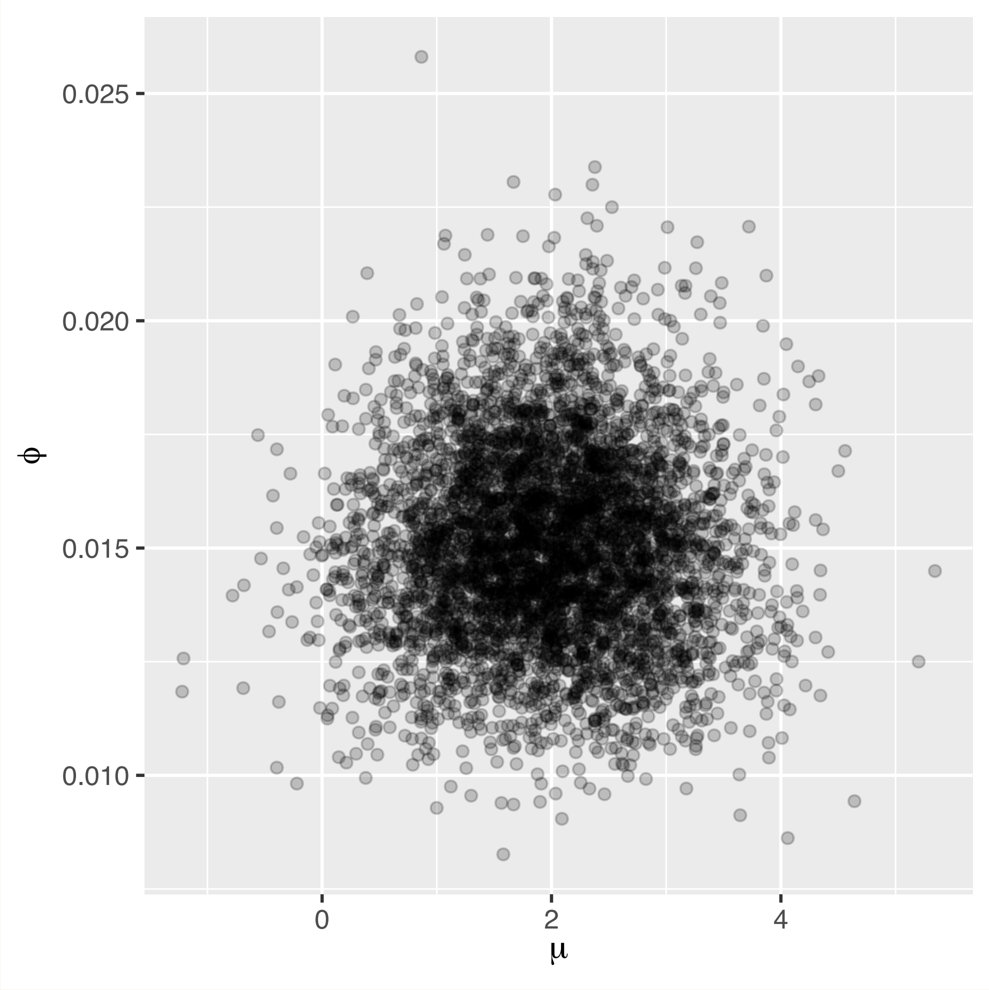
Getting ready to sample

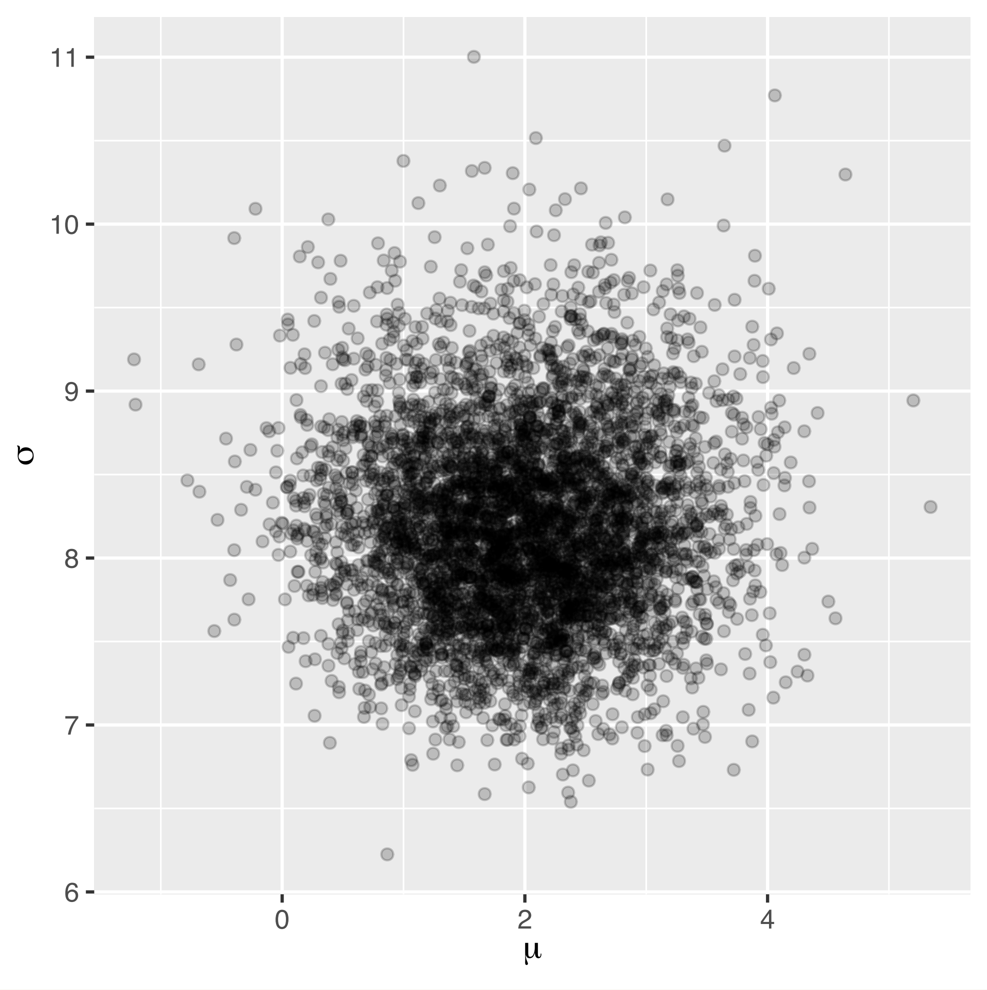
```r
# Data
y <- select_county$pct_change_dem
n <- length(y)

# Prior specification
mu0  <- 0
phi0 <- 1/1000
a    <- 0.1
b    <- 0.1

# Initial parameter values
mu <- mean(y)
s2 <- var(y)
phi <- 1 / s2


# Create empty S x p matrix for MCMC draws
S                    <- 5000
mcmc_draws           <- matrix(NA, nrow = S, ncol = 2)
colnames(mcmc_draws) <- c("mu", "phi")
```

Gibbs sampler

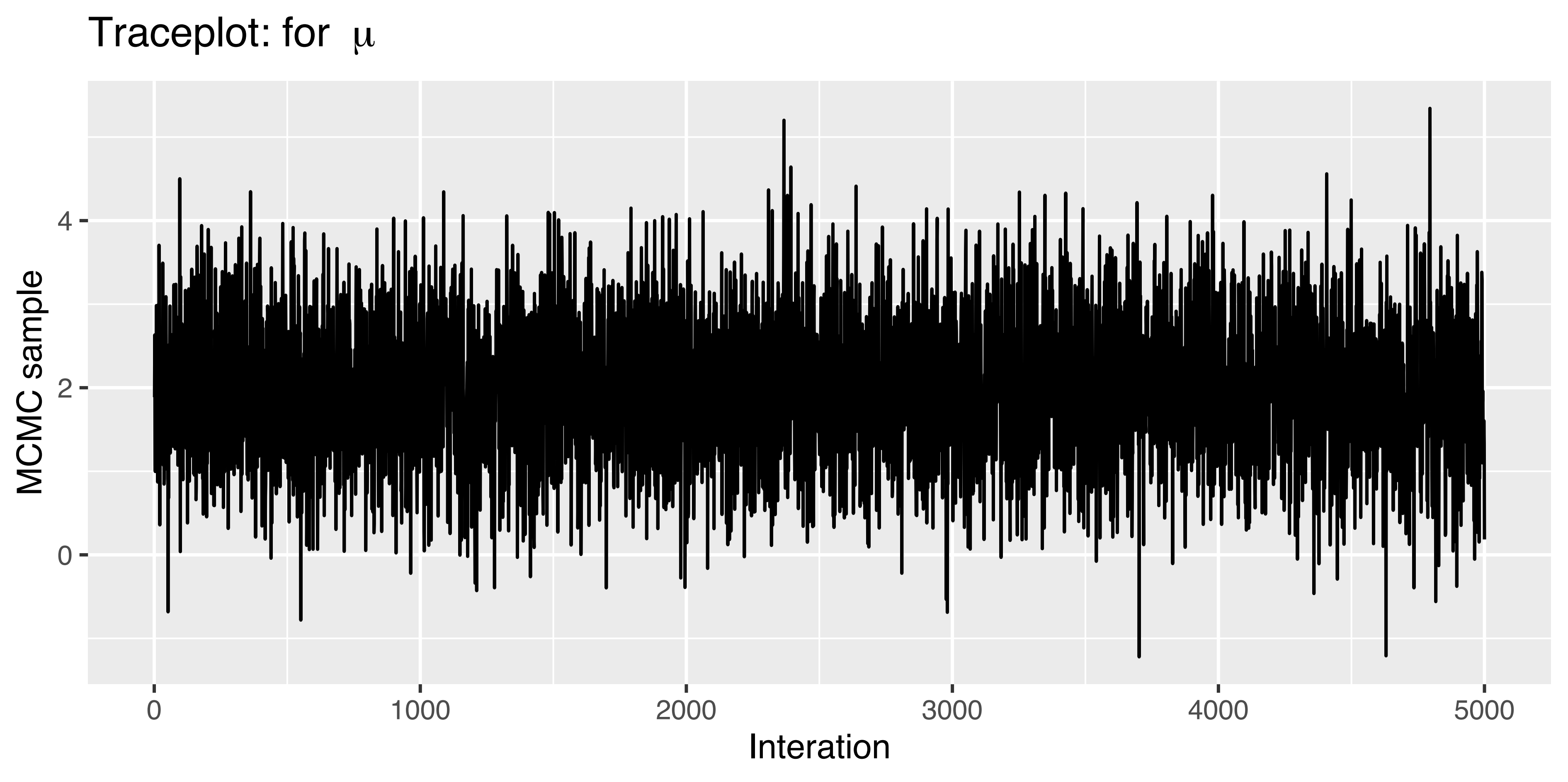
```r
for(i in 1:S) {

  # sample from mu | s2, y
  A    <- sum(y) * phi + mu0 * phi0
  B    <- n * phi + 1 * phi0
  mu   <- rnorm(1, A/B, 1/sqrt(B))

  # sample from s2 | mu, y
  shape  <- n / 2 + a
  scale  <- (sum((y - mu)^2) / 2) + b
  phi    <- rgamma(1, shape, scale)

  # Store the draws
  mcmc_draws[i, ] <- c(mu, phi)
}
```

To get the joint posterior of interest, $\pi(\mu, \sigma | y_1, \ldots, y_n)$, transform ϕ

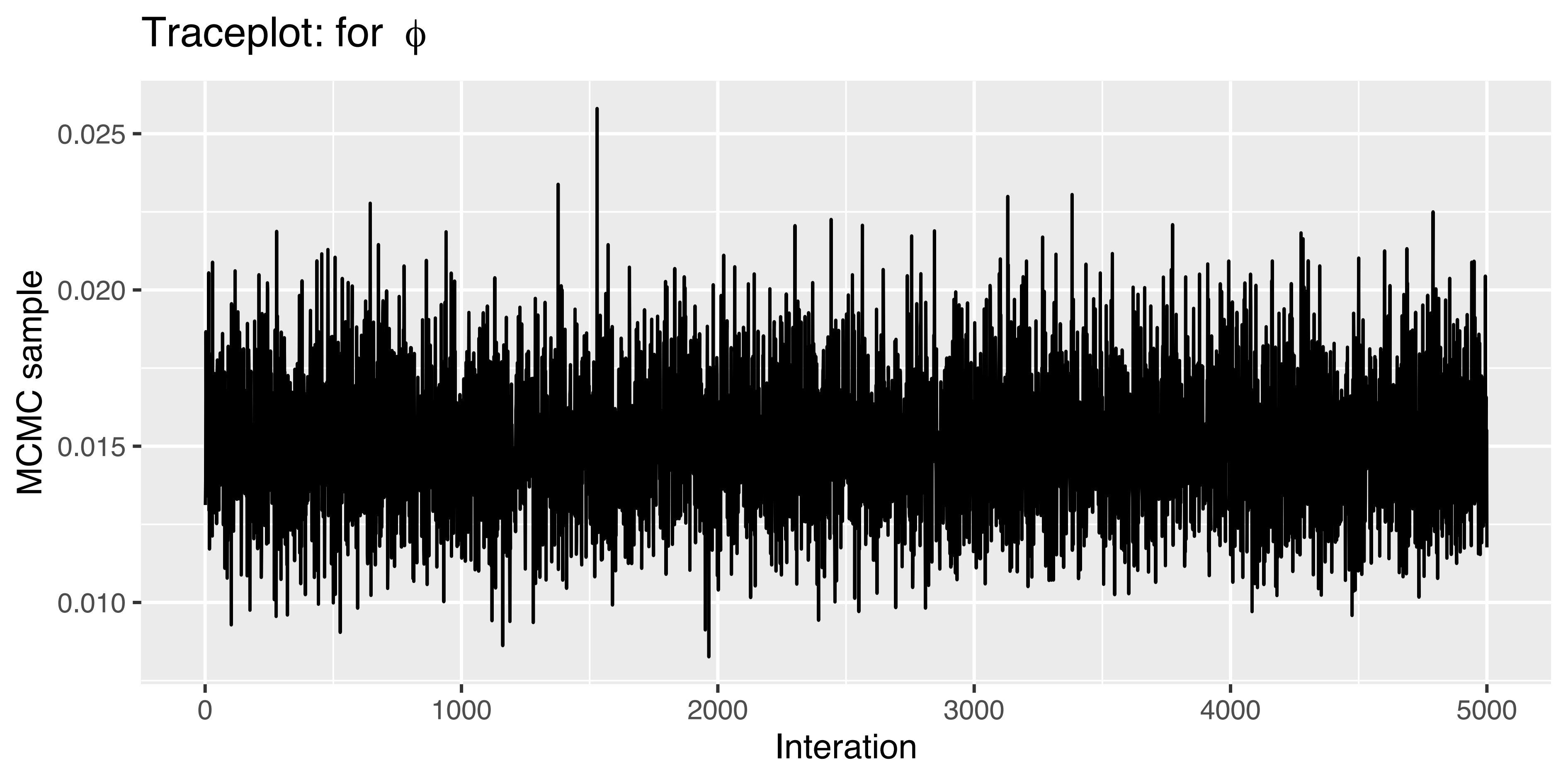




Did the chain converge?



Did the chain converge?

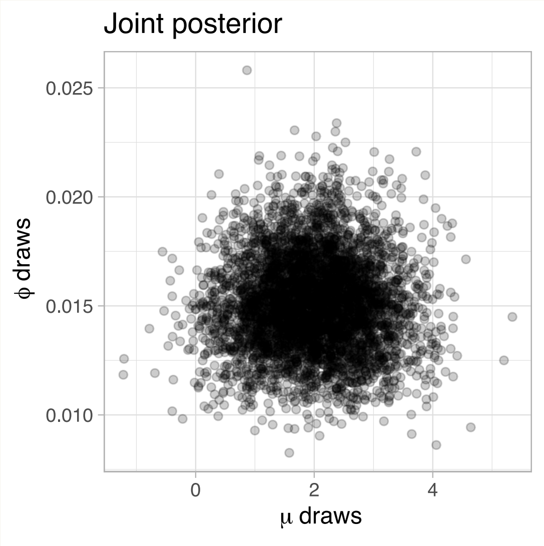


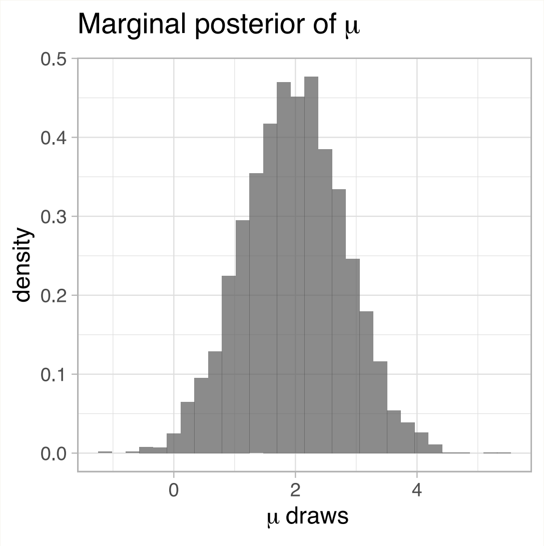
Posterior analysis

Toss out samples prior to convergence (this is called the *burn in* period)

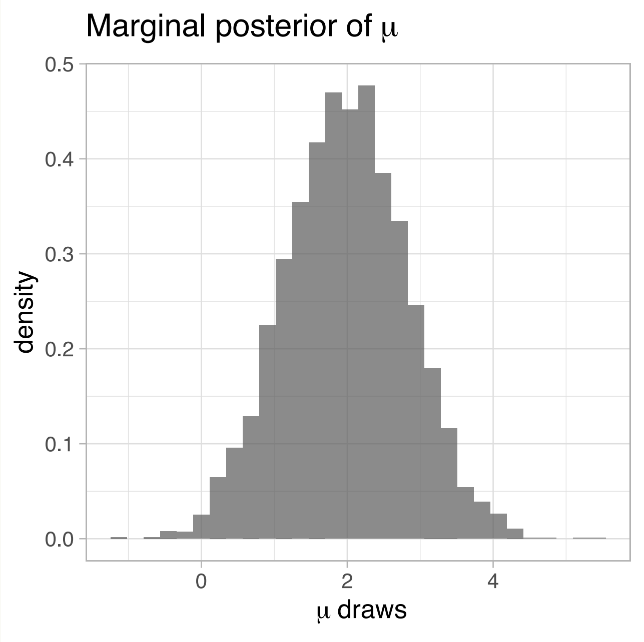
Draw inferences using the remaining MCMC samples just like we have all term

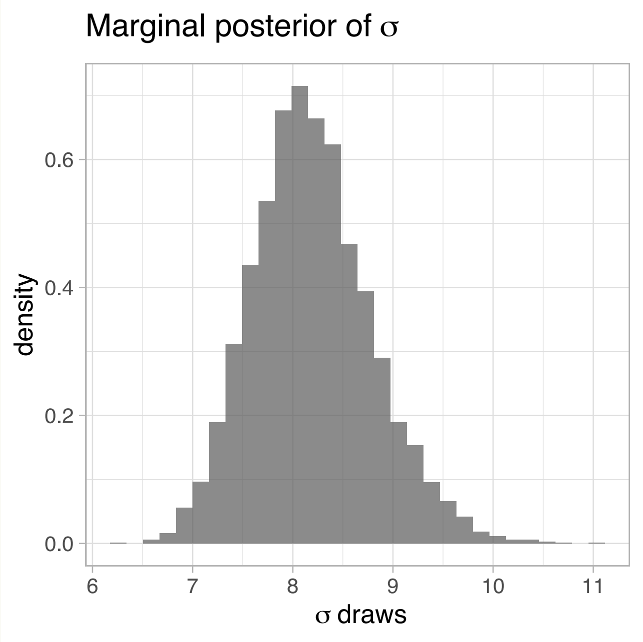
```
no_burn <- mcmc_draws[-c(1:100),]
```





Posterior analysis





	Mean	SD	Q025	Q975
mu	1.98	0.83	0.35	3.61
sigma	8.19	0.59	7.13	9.48

p-stage Gibbs sampler

Target: samples from $\pi(\theta_1, \theta_2, \ldots, \theta_p | y_1, \ldots, y_n)$

1. Set initial values for parameter values, $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)$

2. Draw $\theta_1^{(1)}$ from $\pi(\theta_1 | \theta_2, \ldots, \theta_p, y_1, \ldots, y_n)$

3. Draw $\theta_2^{(1)}$ from $\pi(\theta_2 | \theta_1, \theta_3, \ldots, \theta_p, y_1, \ldots, y_n)$

$\vdots$

p. Draw $\theta_2^{(1)}$ from $\pi(\theta_p | \theta_1, \ldots, \theta_{p-1}, y_1, \ldots, y_n)$

Repeat steps *2-p* S times