BAYESIAN STATISTICS

CLASS 2

WHAT DID WE LEARN FROM CLASS 1?

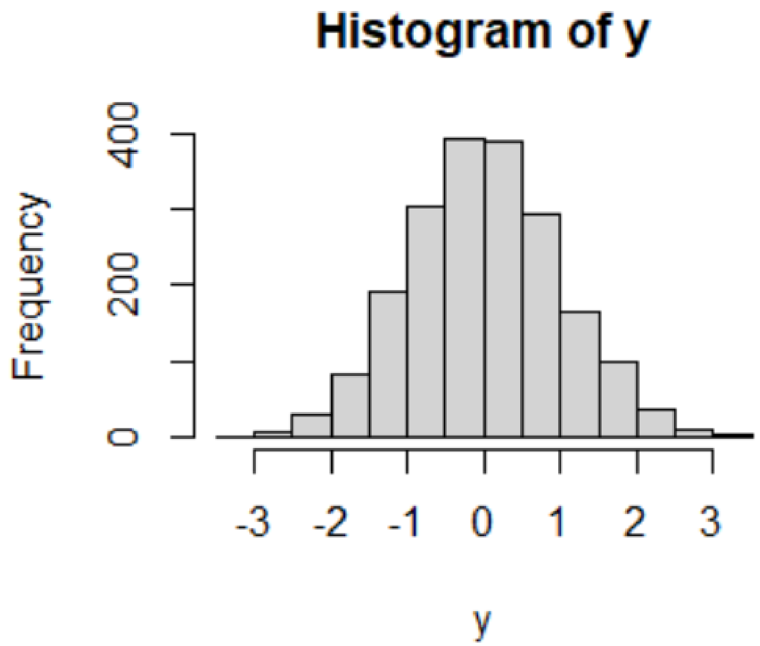
- Terminology: Prior, Sampling distribution, Posterior
- How to define problem (decide sampling distribution of data, define priors for parameters, use Stan to generate posterior distribution of parameters)
- How to use posterior to answer questions about the parameter
- How data (sample size) and prior contribute to the posterior
- Why prior is VERY important when sample size is small

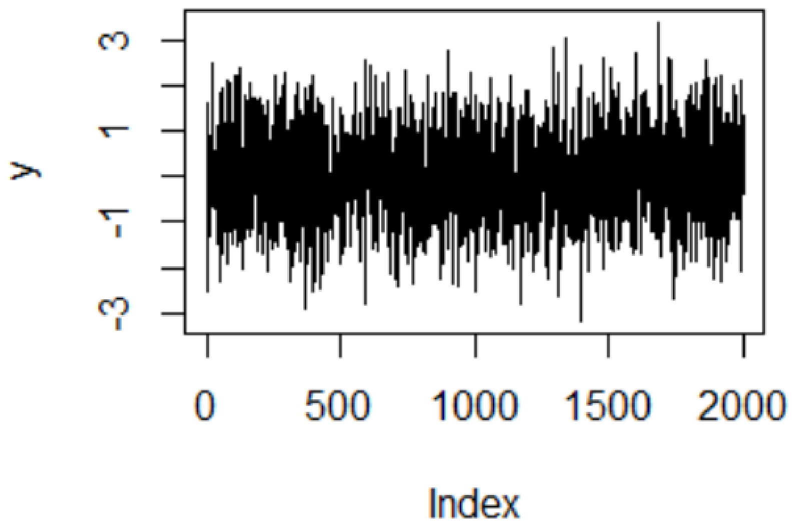
GOALS FOR TODAY

- MCMC - Markov Chain Monte Carlo
  - What it is
  - Has it converged
  - Options to help convergence
- Options in running MCMC to get posterior distribution
- Another in-class example

SIMULATING A DISTRIBUTION

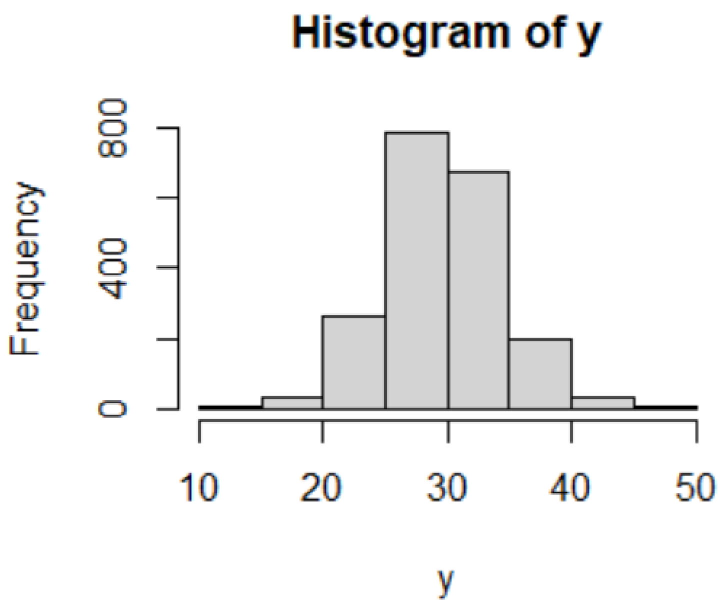
```
y=rnorm(2000)
> hist(y)
> plot(y,type='l')
```

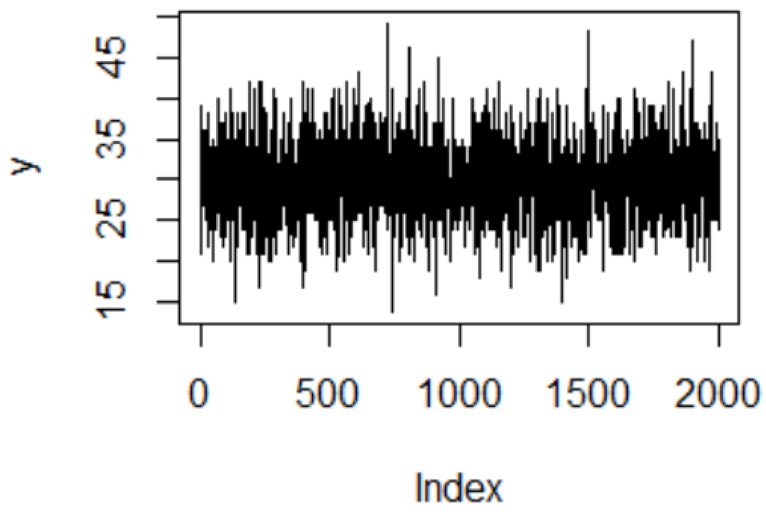




ANOTHER DISTRIBUTION

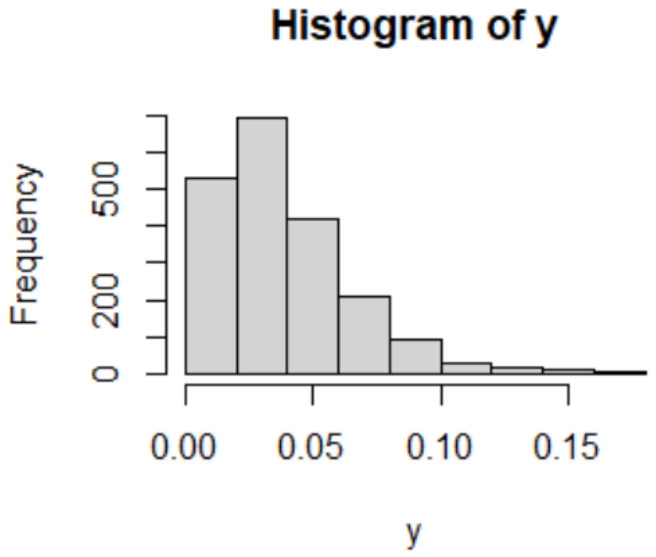
```
> y=rbinom(2000,100,0.3)
> hist(y)
> plot(y,type='l')
```

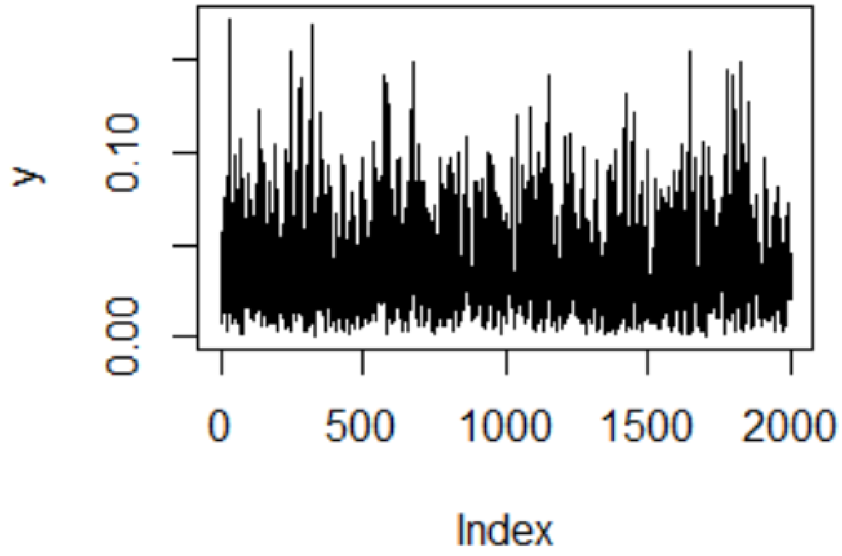




SKEWED DISTRIBUTION

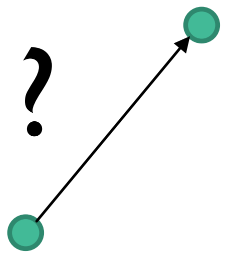
```
> y=rbeta(2000,2,50)
> hist(y)
> plot(y,type='l')
```



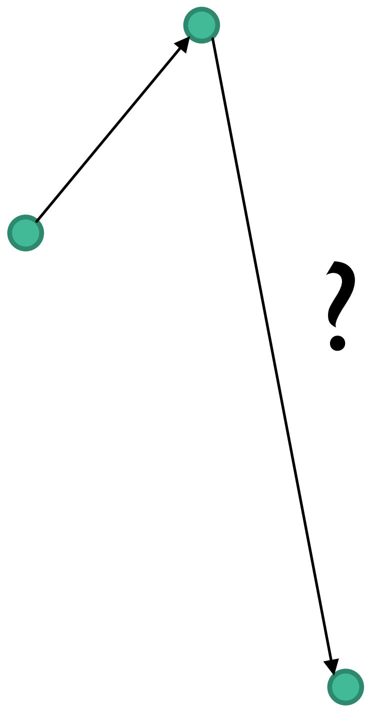


MARKOV CHAIN MONTE CARLO

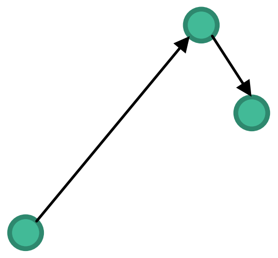
MARKOV CHAIN MONTE CARLO



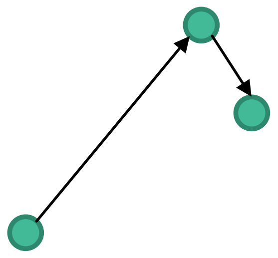
MARKOV CHAIN MONTE CARLO



MARKOV CHAIN MONTE CARLO

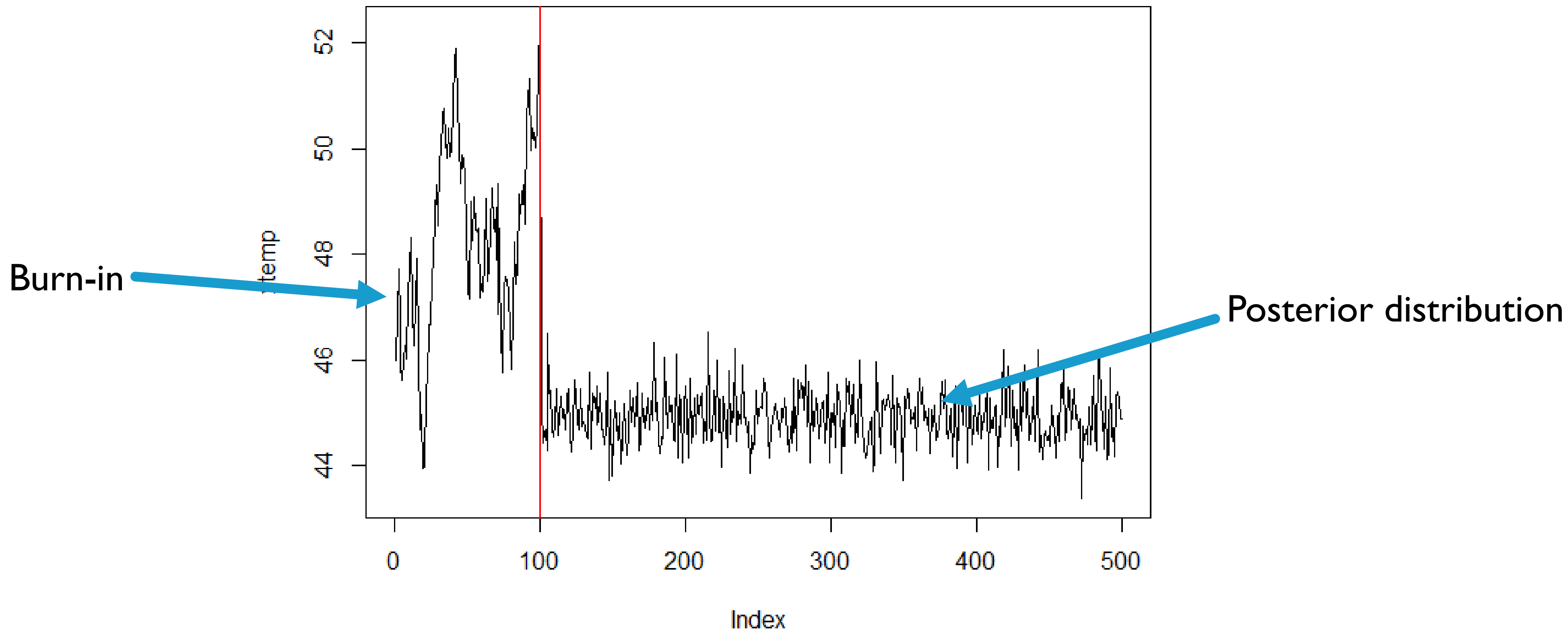


MARKOV CHAIN MONTE CARLO

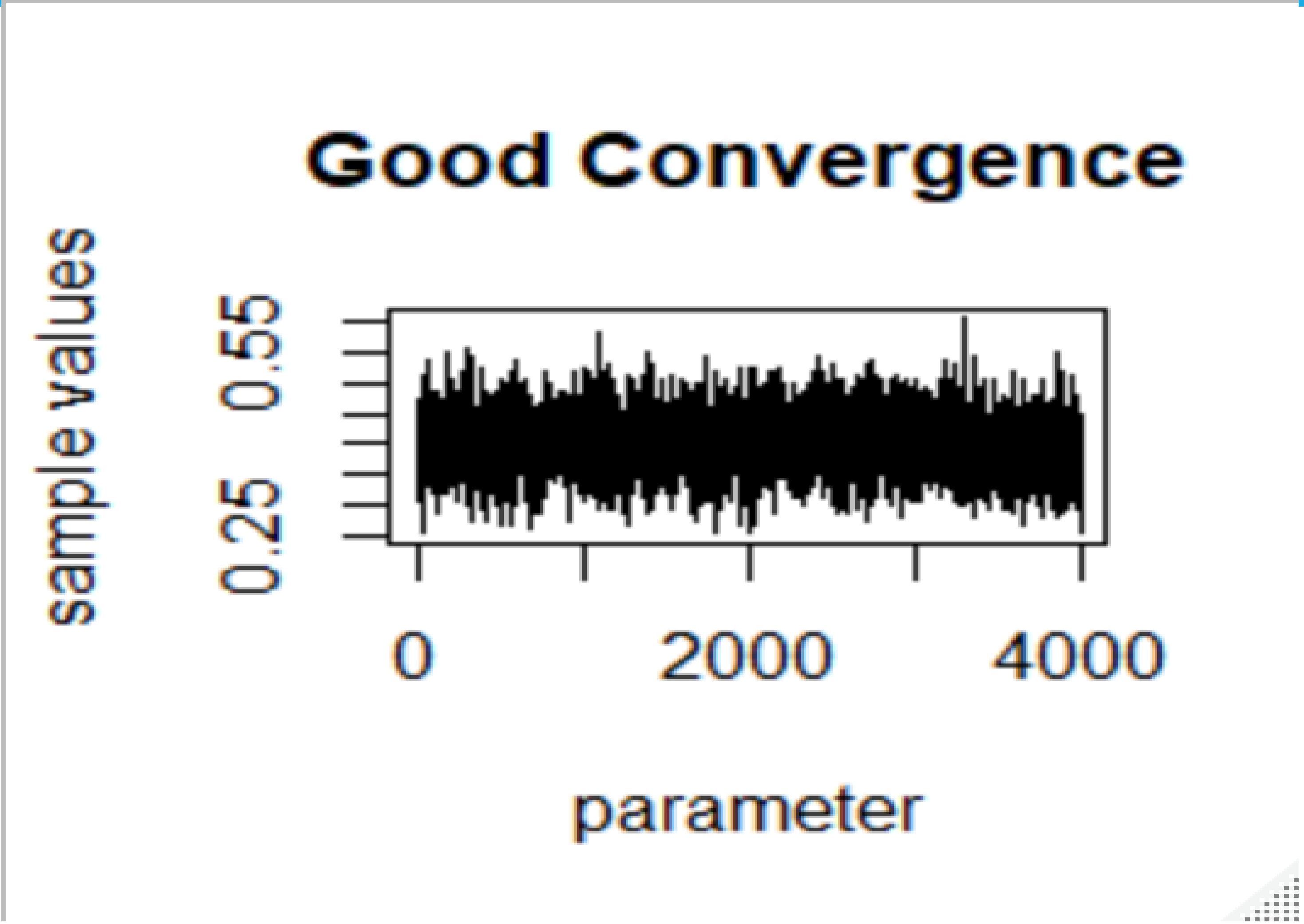


Stan uses the Hamiltonian Monte Carlo method for its Markov Chain and its adaptive variant the no U-turn sampler (NUTS). For more details, see https://mc-stan.org/docs/2_19/reference-manual/hamiltonian-monte-carlo.html

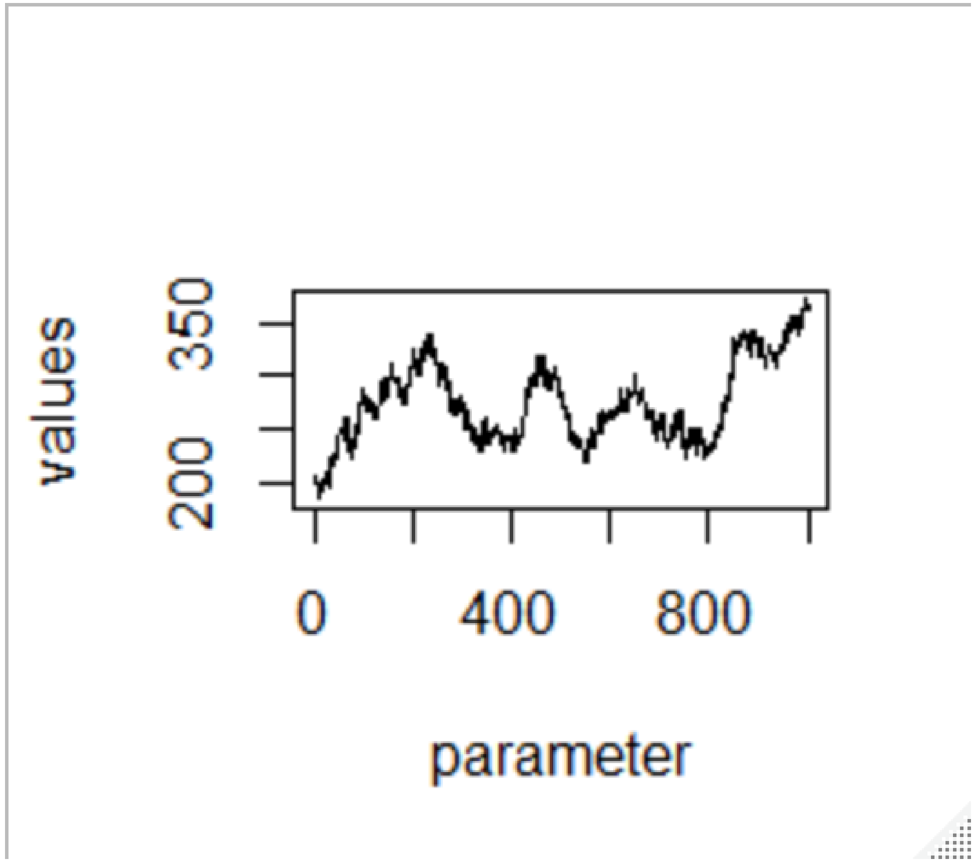
MCMC

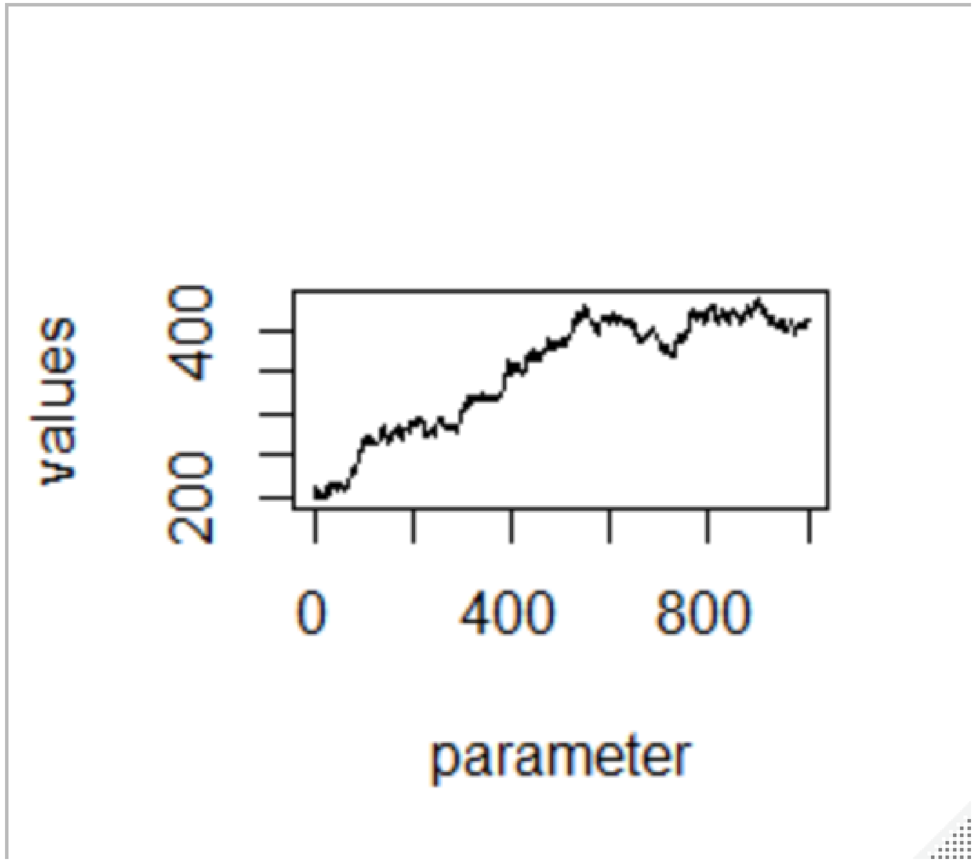

Burn-in
Posterior distribution
temp
52
50
48
46
44
0
100
200
300
400
500
Index

CONVERGENCE



NONCONVERGENCE





FIXES

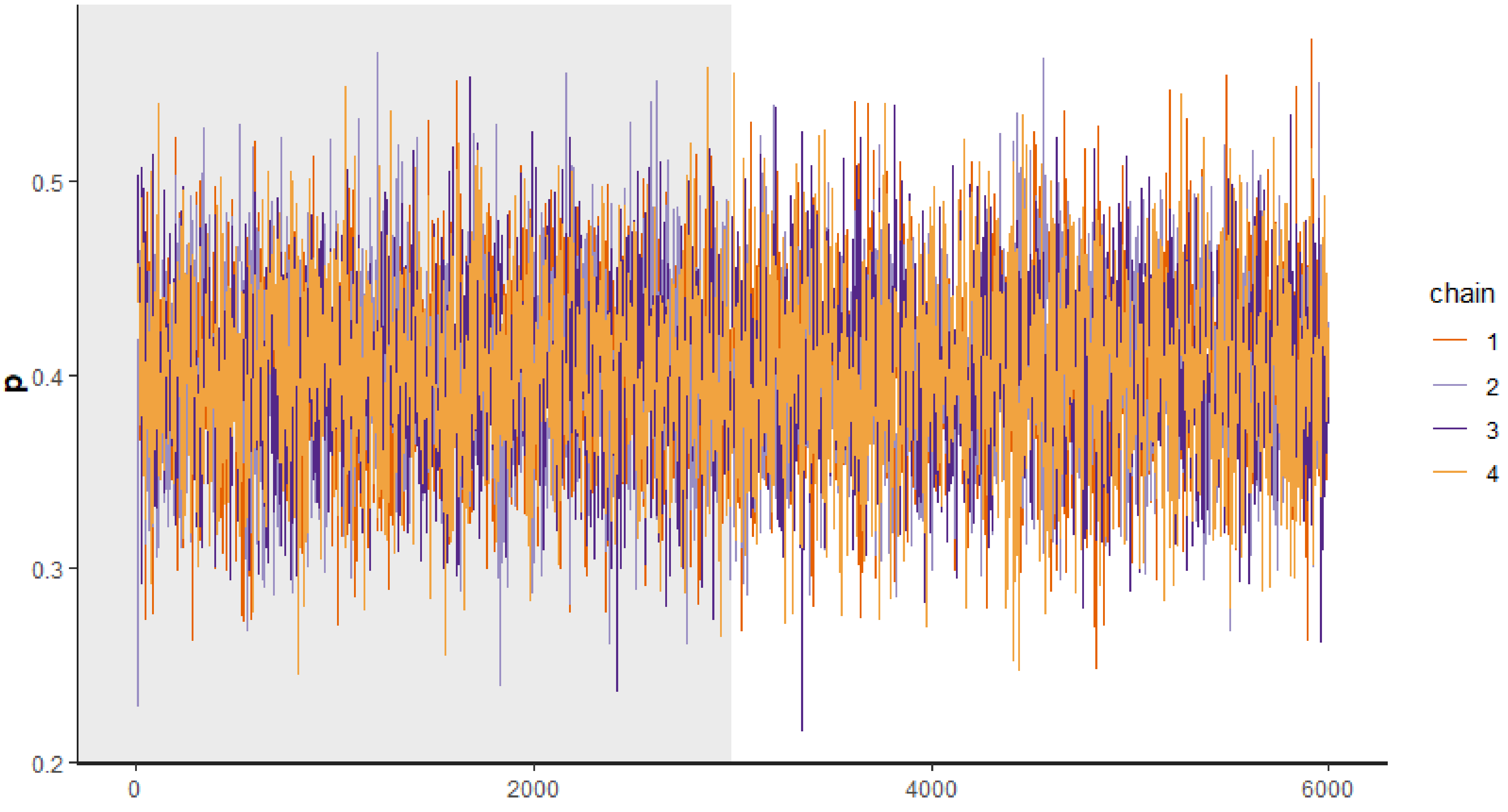
- Improper posterior or bad prior
  - Fix: New prior distribution
- Hasn't converged yet
  - Let the chain run longer
- Chain continues to increase
  - Potentially a bad starting point…provide a new starting point (or change prior)
- Too much autocorrelation in chain
  - Thin the chain

OPTIONS FOR STAN CODE

```
binom.stan=stan(file='Q:\\My Drive\\Bayesian\\Code\\Binomial_example.stan',data=binom.data,
chains = 4,           # number of Markov chains
  warmup = 3000,   # number of warmup iterations per chain
  iter = 6000,         # total number of iterations per chain
  refresh = 0,          # no progress shown
  thin=3,                 # will 'thin' the chains…help with autocorrelated posterior samples
  init=0.3                #specify initial values..I only did it for one chain
  )
```

Creates four chains; each chain has 6000 values, however, only every 3rd value is taken (now down to 2000 per chain that is useful); first 3000 (well, actually only 1000 since we are thinning) is burn-in meaning it is not used

End result will have a total of 4000 posterior values (1000 from each chain) traceplot(binom.stan, inc_warmup = TRUE)



mean se_mean sd 10% 50% 90% n_eff Rhat p 0.40 0.00 0.05 0.34 0.40 0.46 3268 1 lp__ -69.23 0.01 0.72 -70.10 -68.96 -68.74 3290 1

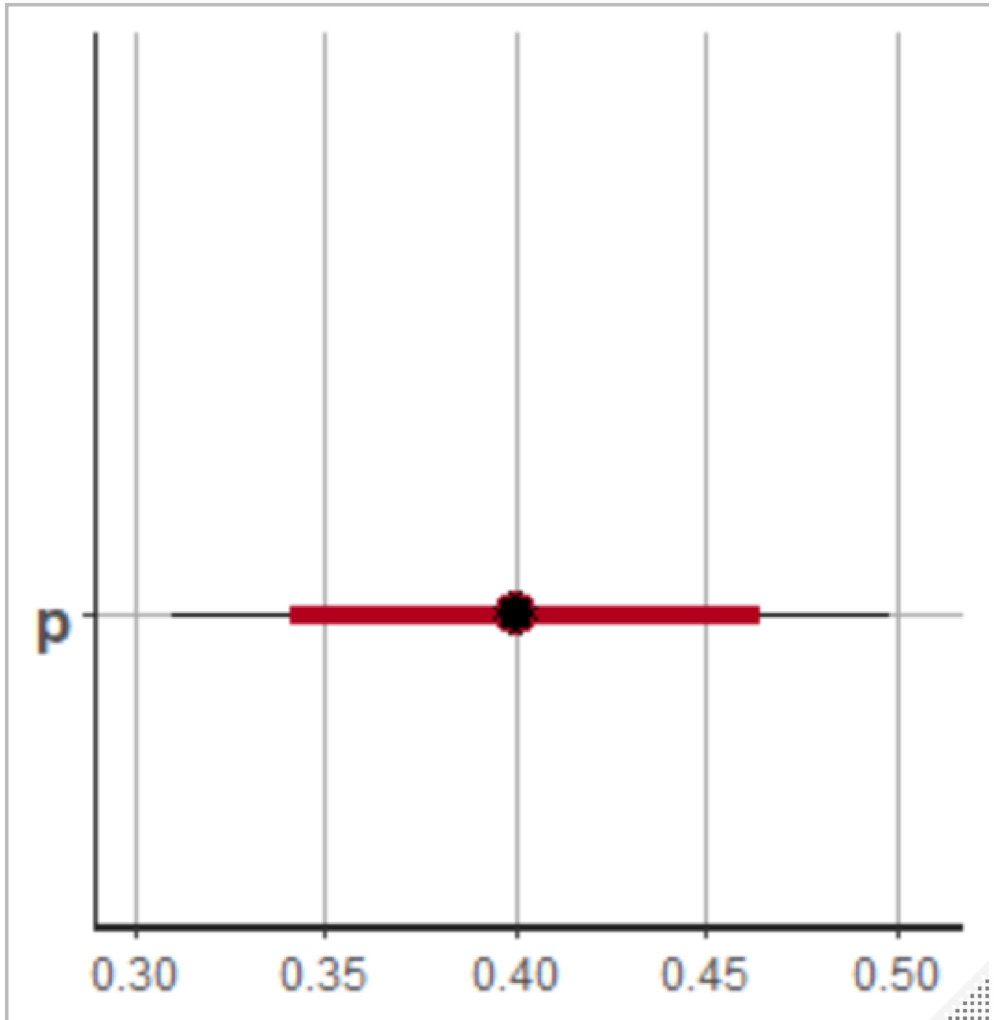
DIAGNOSTICS

print(binom.stan, pars=c("p", "lp__"), probs=c(.1,.5,.9))

	mean	se_mean	sd	10%	50%	90%	n_eff	Rhat
p	0.40	0.00	0.05	0.34	0.40	0.46	3268	1
lp__	-69.23	0.01	0.72	-70.10	-68.96	-68.74	3290	1

Log posterior value (unnormalized)

Want these to be close to 1 (this means convergence); if greater than 1.1, then there could potentially be a problem (Rhat or potential scale reduction) plot(binom.stan)



Inner is 80% probability interval
Outer is 95% probability interval

ANOTHER POTENTIAL WARNING

Warning: There were 2 divergent transitions after warmup. Increasing adapt_delta above 0.8 may help. See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

Fix: add this line to your options…

```
control=list(adapt_delta=0.9)
```

EMPIRICAL BAYES

- Empirical Bayes is between Full Bayesian Statistics and frequentist
- It uses data to estimate the prior distribution parameters
- However, if you ‘double-dip’ with the data, you are adding extra variability and this needs to be accounted for in the in the posterior distribution (be cautious of this!!)

NEW EXAMPLE

- WARP BREAKS per loom
- Would like to understand the mean number of warp breaks per loom (a loom corresponds to a fixed length of yarn)

Sampling distribution:

Number of parameters:

Prior distribution: