

# Introduction to Bayesian Statistics

---

Shin-ichi Mayekawa

[mayekawa@hum.titech.ac.jp](mailto:mayekawa@hum.titech.ac.jp)

[www.ms.hum.titech.ac.jp](http://www.ms.hum.titech.ac.jp)

# Gibbs Sampler

- Gibbs Sampler

— 1 : Two Variate Gibbs sample —

1. Estimate  $f^0(Y)$
2. Draw  $Y^{(0)} = y^{(0)}$  according to  $f^0(Y)$ .
3. Draw  $X^{(1)} = x^{(1)}$  according to  $f_{X|Y}(X|Y = y^{(0)})$
4. Draw  $Y^{(1)} = y^{(1)}$  according to  $f_{Y|X}(Y|X = x^{(1)})$
5. Draw  $X^{(2)} = x^{(2)}$  according to  $f_{X|Y}(X|Y = y^{(1)})$
6. Draw  $Y^{(2)} = y^{(2)}$  according to  $f_{Y|X}(Y|X = x^{(2)})$
- $\vdots$      $\vdots$
7. Draw  $X^{(t+1)} = x^{(t+1)}$  according to  $f_{X|Y}(X|Y = y^{(t)})$
8. Draw  $Y^{(t+1)} = y^{(t+1)}$  according to  $f_{Y|X}(Y|X = x^{(t+1)})$
- $\vdots$      $\vdots$

# Normal Model with Semi Natural Conjugate Prior

## ■ prior

$$\mu \sim N(\gamma_0, \phi_0^2)$$

$$\sigma \sim \chi^{-1}(\kappa_0^2, \nu_0)$$

$$h(\mu, \sigma) = h(\mu)h(\sigma)$$

$$\propto \frac{1}{\sqrt{\phi_0^2}} \exp\left(-\frac{1}{2\phi_0^2}(\mu - \gamma_0)^2\right) \times \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right)$$

## ■ posterior

$$h(\mu, \sigma | \mathbf{y}) \propto f(\mathbf{y} | \mu, \sigma) h(\mu, \sigma)$$

$$\propto \exp\left[-\frac{1}{2} \left( \frac{\text{RSS}_y + \kappa_0^2}{\sigma^2} + \frac{N(\gamma_0 - \bar{y})^2}{\sigma^2 + N\phi_0^2} + \left(\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}\right) \left(\mu - \frac{\gamma_0\sigma^2 + N\phi_0^2\bar{y}}{\sigma^2 + N\phi_0^2}\right)^2 \right)\right]$$

$$\times \sigma^{-N-\nu_0-1} \phi_0^{-1}$$

cannot be identified.

# Normal Model with Semi Natural Conjugate Prior

- full conditional posterior

$$h(\mu, |\sigma, \mathbf{y}) \propto \exp \left[ -\frac{1}{2} \left( \frac{N}{\sigma^2} + \frac{1}{\phi_0^2} \right) \left( \mu - \frac{\gamma_0 \sigma^2 + N \phi_0^2 \bar{y}}{\sigma^2 + N \phi_0^2} \right)^2 \right]$$

$$\mu | \sigma, \mathbf{y} \sim N \left[ \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \left( \frac{1}{\phi_0^2} \gamma_0 + \frac{N}{\sigma^2} \bar{y} \right), \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \right]$$

$$h(\sigma | \mu, \mathbf{y}) \propto \sigma^{-(N+\nu_0)-1} \times \exp \left[ -\frac{1}{2} \left( \frac{\text{RSS}_y + N(\mu - \bar{y})^2 + \kappa_0^2}{\sigma^2} \right) \right]$$

$$\sigma | \mu, \mathbf{y} \sim \chi^{-1} (N + \nu_0, N(\mu - \bar{y})^2 + \text{RSS}_y + \kappa_0^2)$$

# Example of Gibbs Sampler

- prior

$\gamma_0 = 50, \phi_0 = 10, \nu_0 = 2, \kappa_0 = 12.5$  (prior mean  $\sigma = 5$ )

- data

$N = 20, \text{mean} = 100, \text{std} = 10$

# Example of Gibbs Sampler

- generate  $X:\mu$  given  $Y:\sigma$

$$\mu|\sigma, \mathbf{y} \sim N \left[ \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \left( \frac{1}{\phi_0^2} \gamma_0 + \frac{N}{\sigma^2} \bar{y} \right), \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \right]$$

- generate  $Y:\sigma$  given  $X:\mu$

$$\sigma|\mu, \mathbf{y} \sim \chi^{-1}(N + \nu_0, N(\mu - \bar{y})^2 + \text{RSS}_y + \kappa_0^2)$$

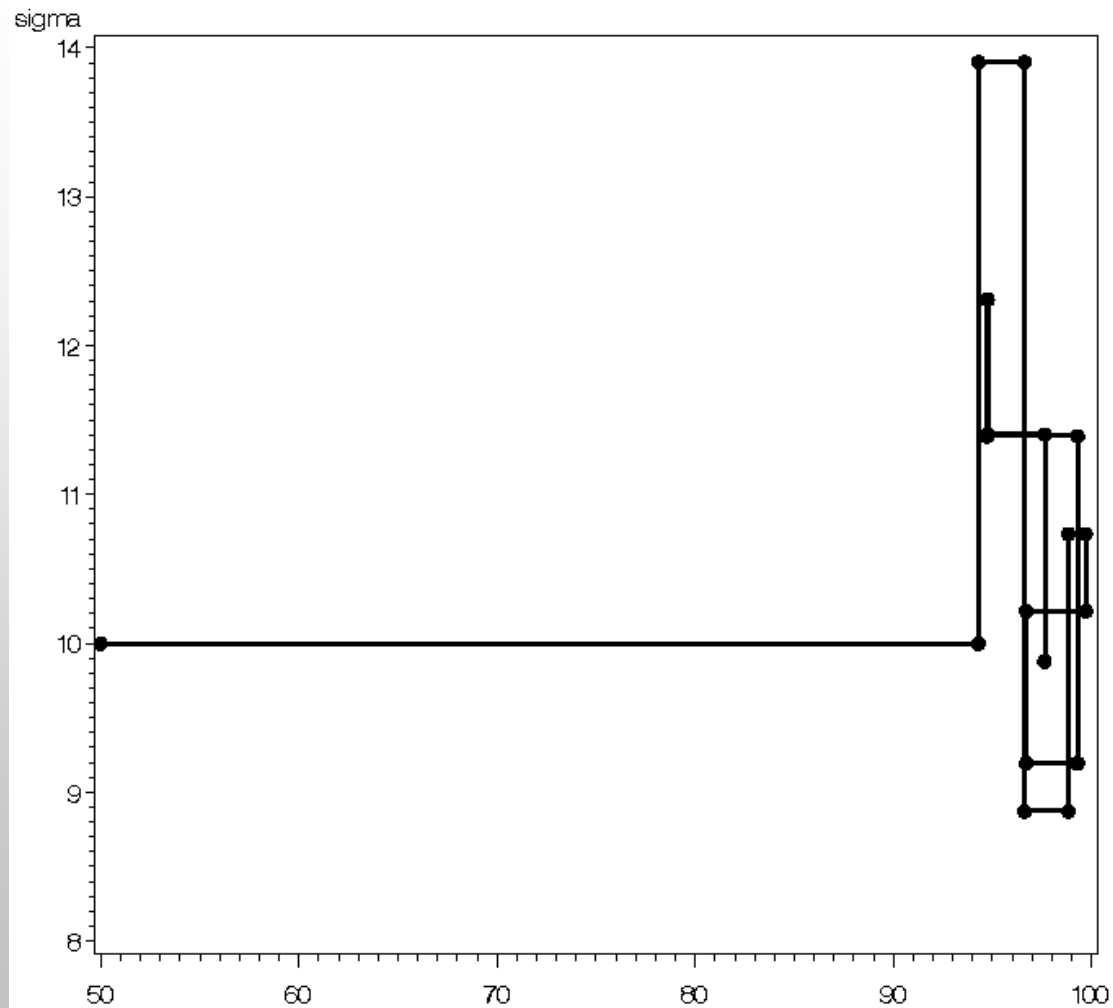
# Example of Gibbs Sampler

- generated random numbers

rep	mu	sigma
1	50.0000	10.0000
2	94.3247	10.0000
2	94.3247	13.9047
3	96.6447	13.9047
3	96.6447	8.8730
4	98.8580	8.8730
4	98.8580	10.7365
5	99.7417	10.7365
5	99.7417	10.2181
6	96.7341	10.2181
6	96.7341	9.1961
7	99.3172	9.1961
7	99.3172	11.3930
8	94.7472	11.3930
8	94.7472	12.3099
9	94.7861	12.3099
9	94.7861	11.4048
10	97.6591	11.4048
10	97.6591	9.8787

# Example of Gibbs Sampler

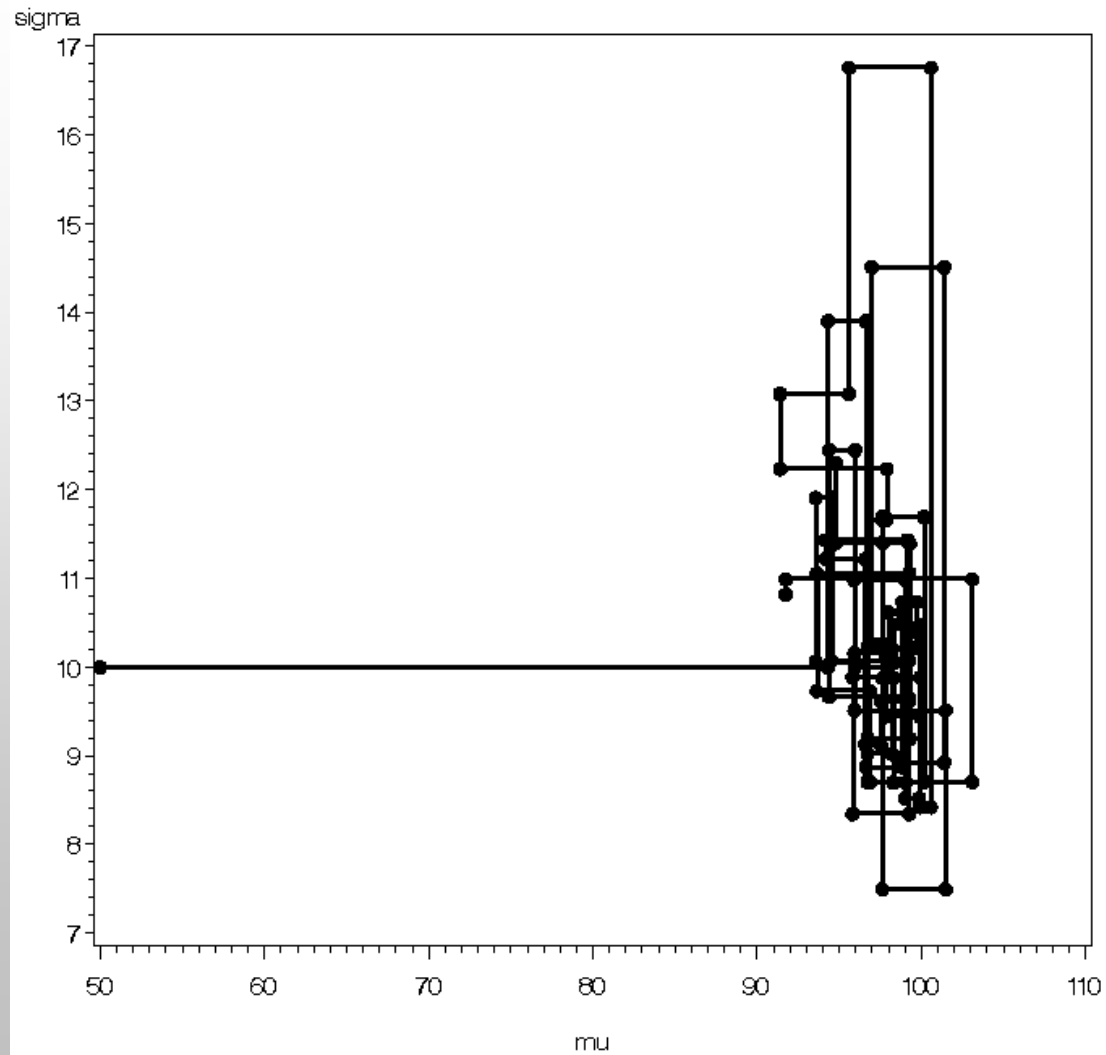
- bivariate history





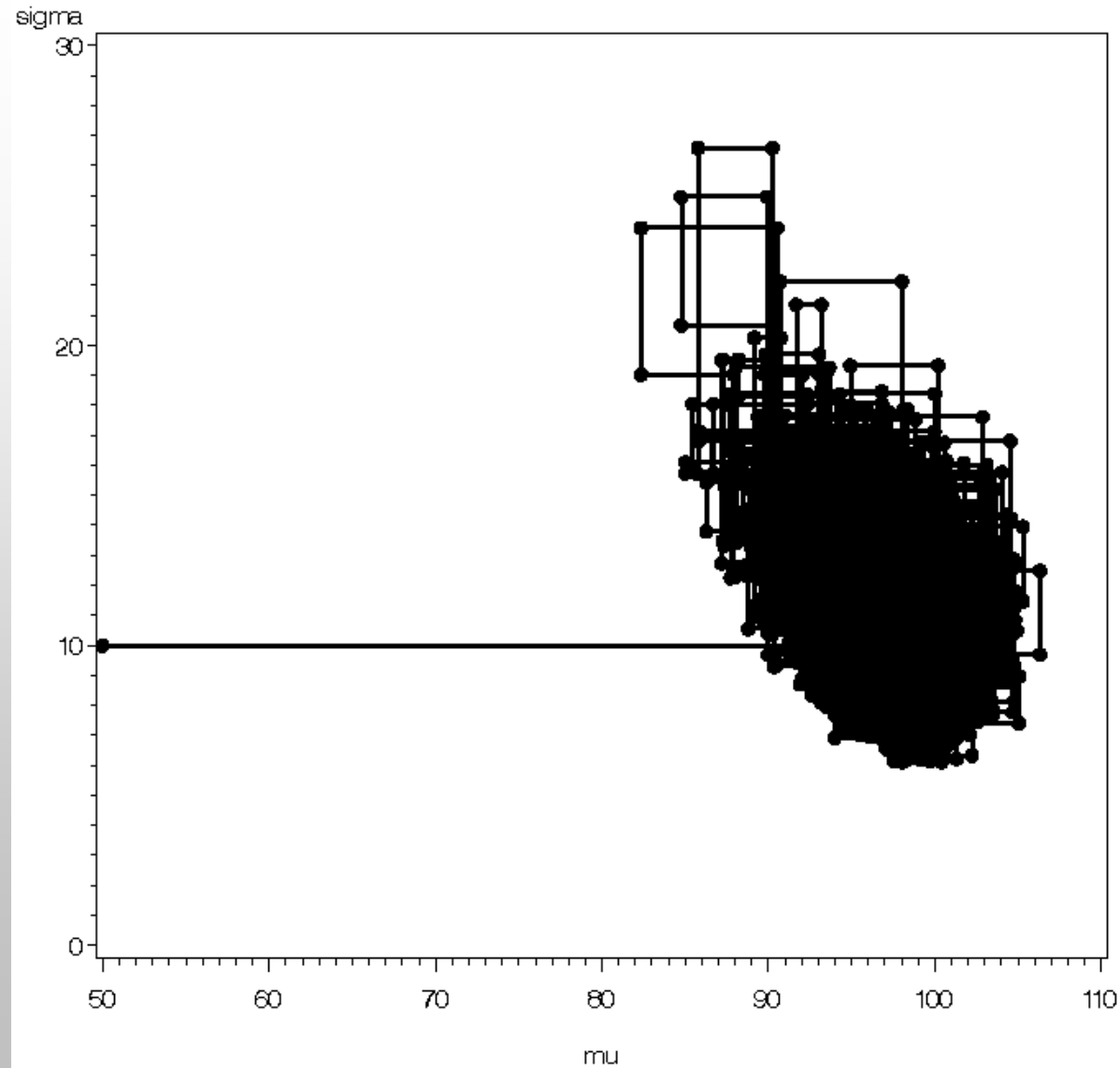
# Example of Gibbs Sampler

- bivariate history



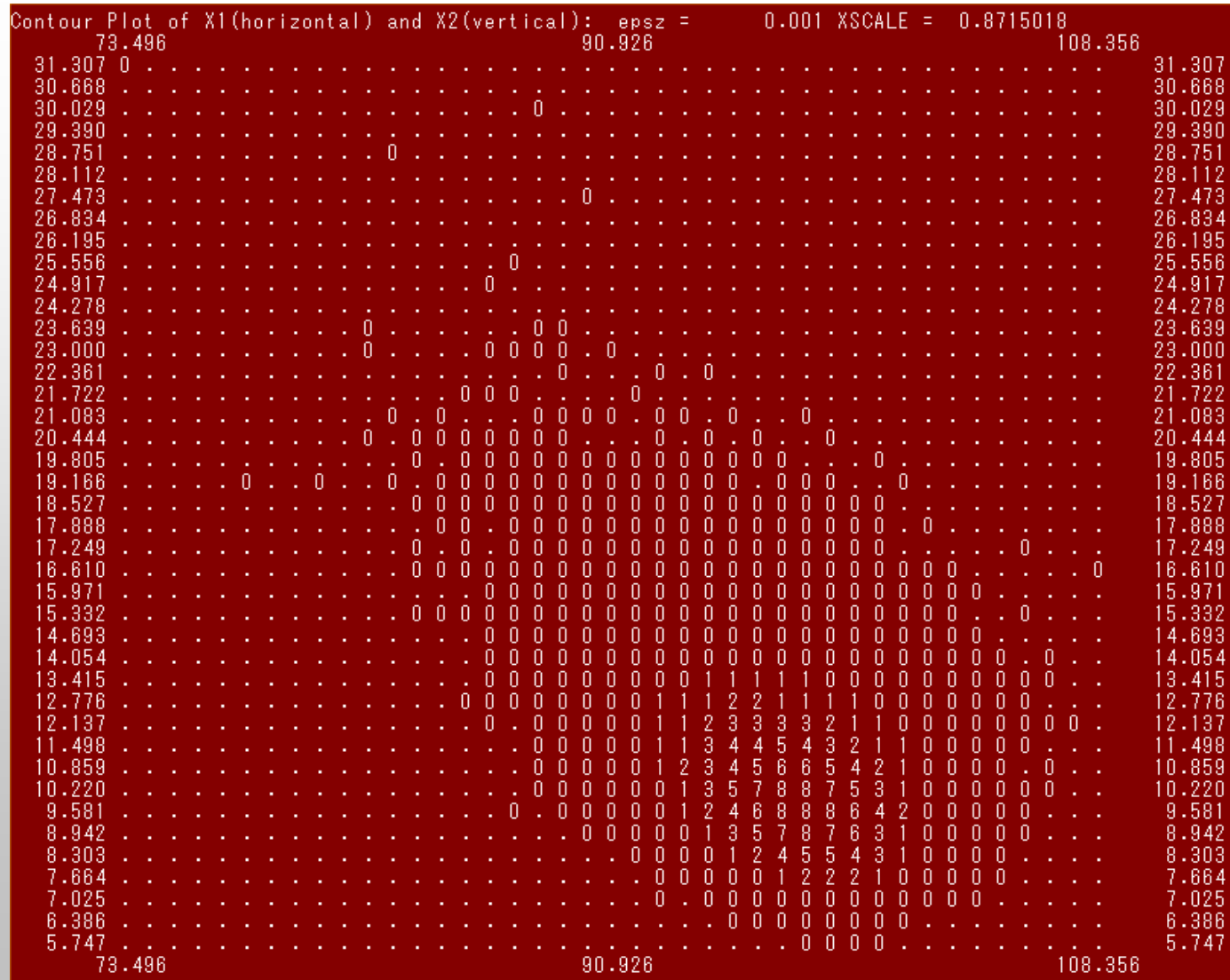
# Example of Gibbs Sampler

- bivariate history



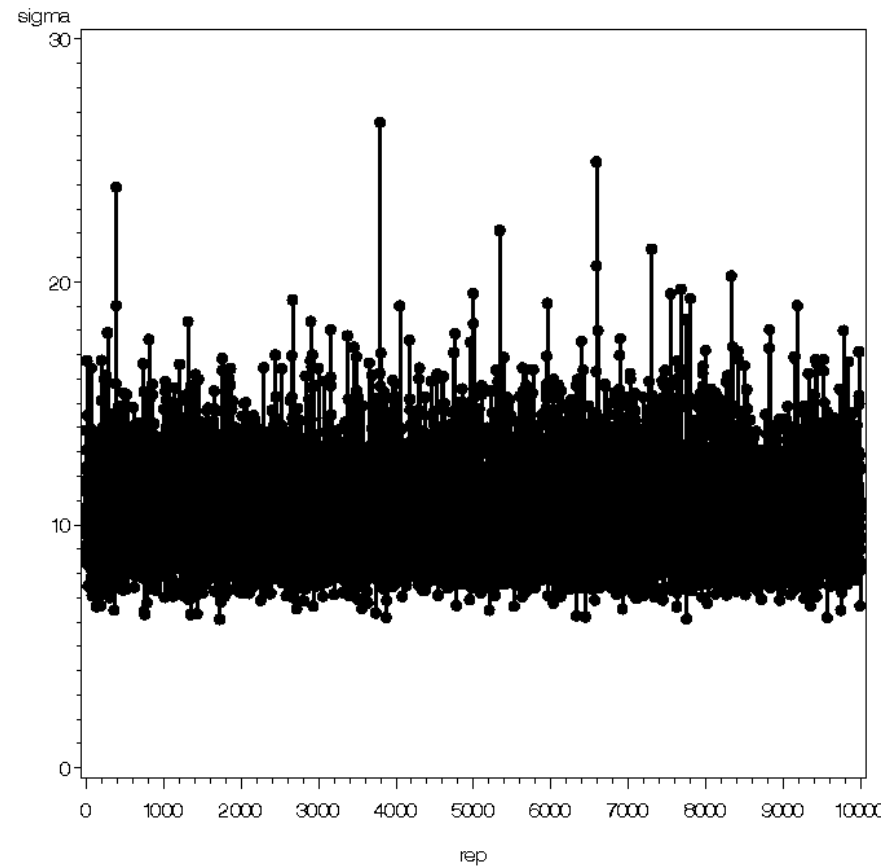
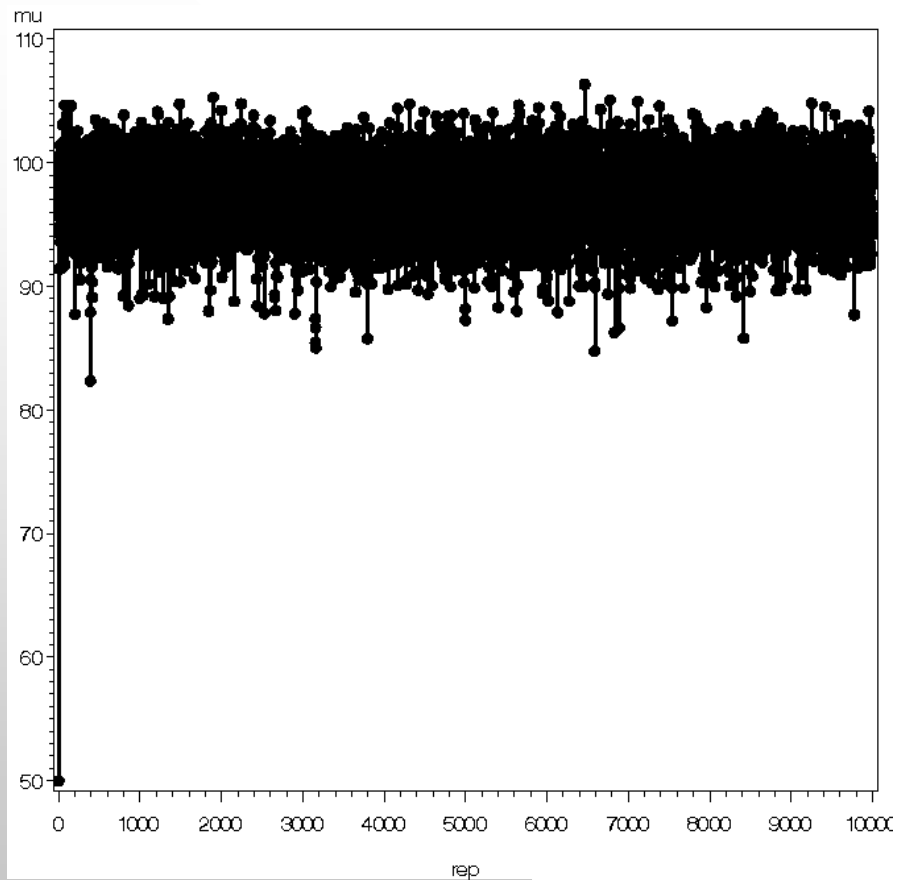
# Example of Gibbs Sampler

- joint distribution



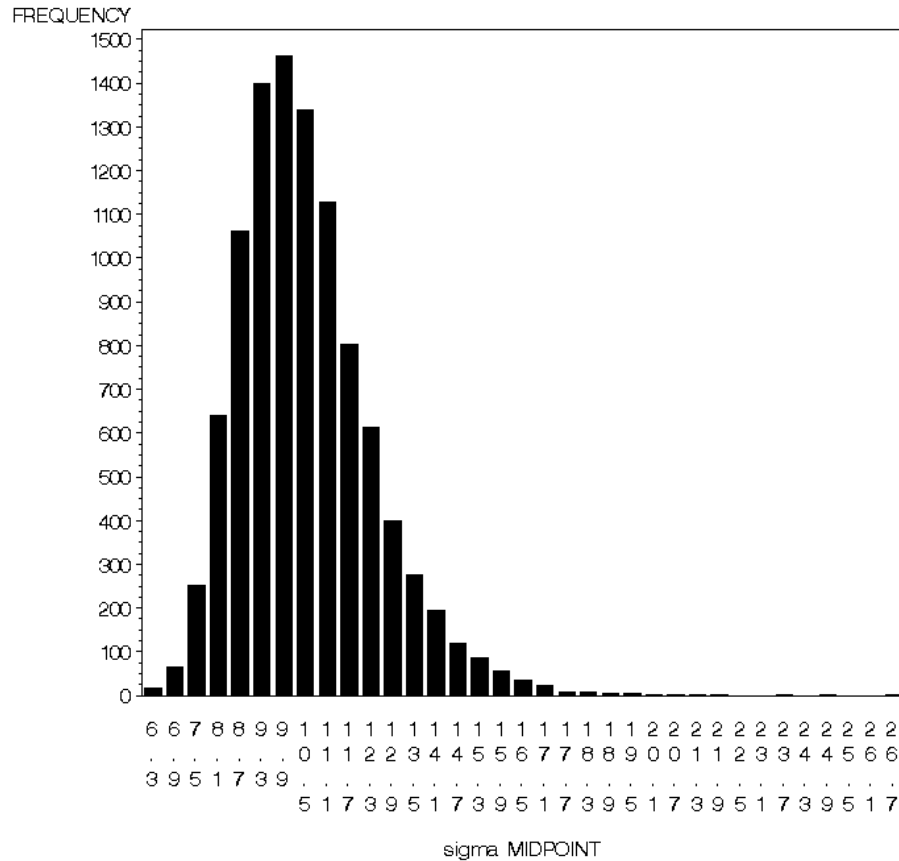
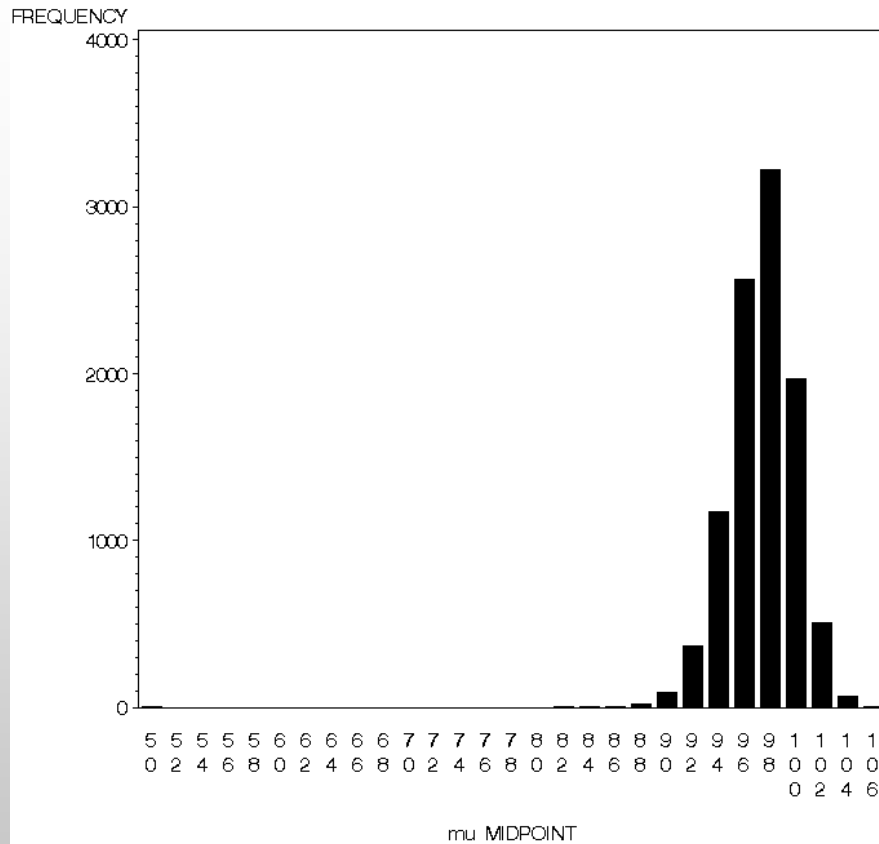
# Example of Gibbs Sampler

- marginal history plot



# Example of Gibbs Sampler

- marginal distribution of random numbers



# Semi Natural Conjugate Prior

```
model
{
  # prior distribution of (beta, sigma)
  # prior sigma distributed as invchi( nu, kappa2 )
  c2 ~ dchisqr( nu )      # chi2( nu )
  t2 <- c2/kappa2        # chi2( nu, 1/kapp2 )
  sigma <- 1/sqrt(t2)    # invchi( nu, kappa2 )

  # prior beta
  beta ~ dnorm( gamma, Phiinv )

  # likelihood
  for( i in 1:nobs ) {
    y[i] ~ dnorm(beta, t2)
  }
}
```

# Semi Natural Conjugate Prior

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
beta	97.27	2.510	0.02510	0.02951
sigma	10.55	1.895	0.01895	0.02195

2. Quantiles for each variable:

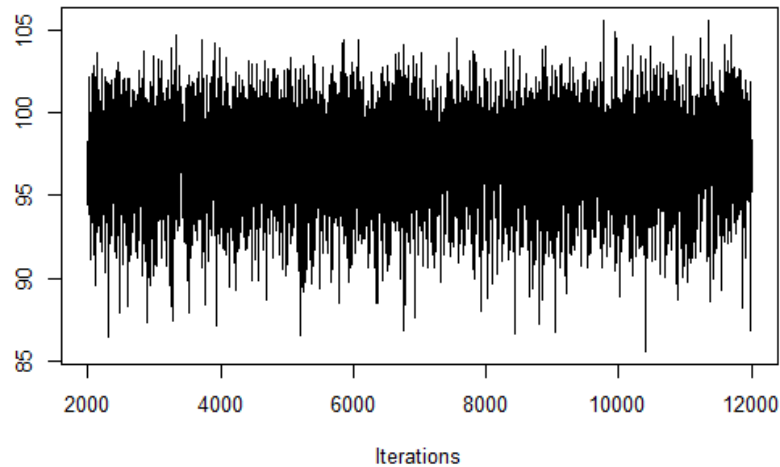
	2.5%	25%	50%	75%	97.5%
beta	91.820	95.724	97.38	99.01	101.88
sigma	7.688	9.229	10.25	11.55	14.97

correlations

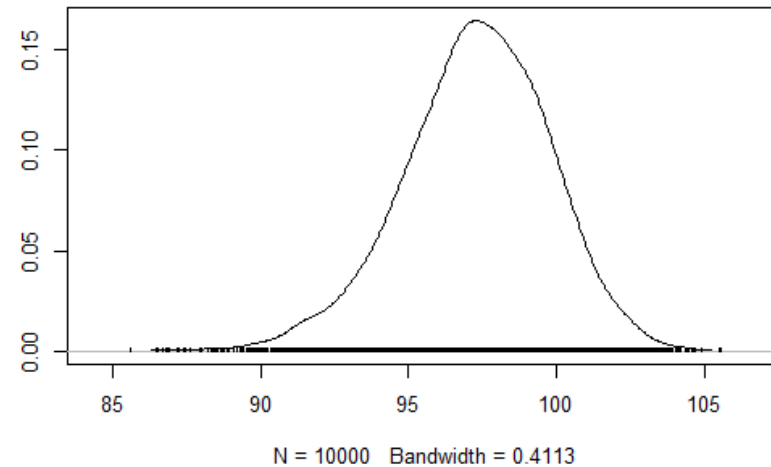
	beta	sigma
beta	1.0000000	-0.3968317
sigma	-0.3968317	1.0000000

# Semi Natural Conjugate Prior

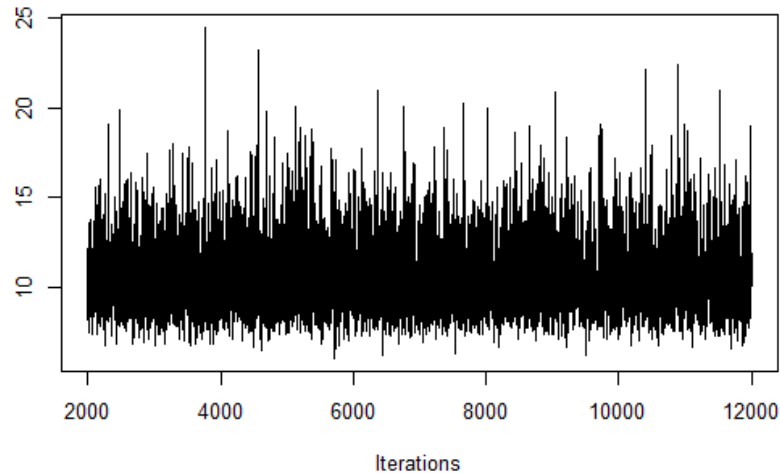
Trace of beta



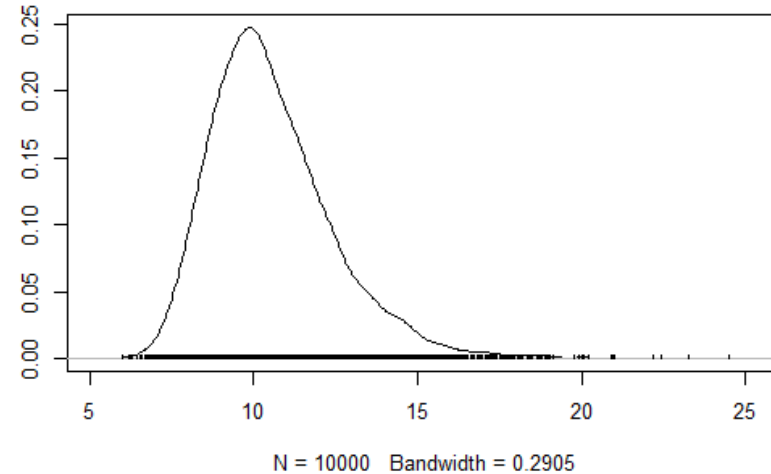
Density of beta



Trace of sigma



Density of sigma





# Semi Natural Conjugate Prior



# Natural Conjugate Prior

```
model
{
  # prior distribution of (beta, sigma)
  # prior sigma distributed as invchi( nu, kappa2 )
  c2 ~ dchisqr( nu )      # chi2( nu )
  t2 <- c2/kappa2        # chi2( nu, 1/kappa2 )
  sigma <- 1/sqrt(t2)    # invchi( nu, kappa2 )

  # conditional prior precision Phiinv2 of beta given sigma (t2)
  Phiinv2 <- Phiinv*t2
  # prior beta
  beta ~ dnorm( gamma, Phiinv2 )

  # likelihood
  for( i in 1:nobs ) {
    y[i] ~ dnorm(beta, t2)
  }
}
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
beta	99.942	2.261	0.02261	0.02261
sigma	9.978	1.589	0.01589	0.01703

2. Quantiles for each variable:

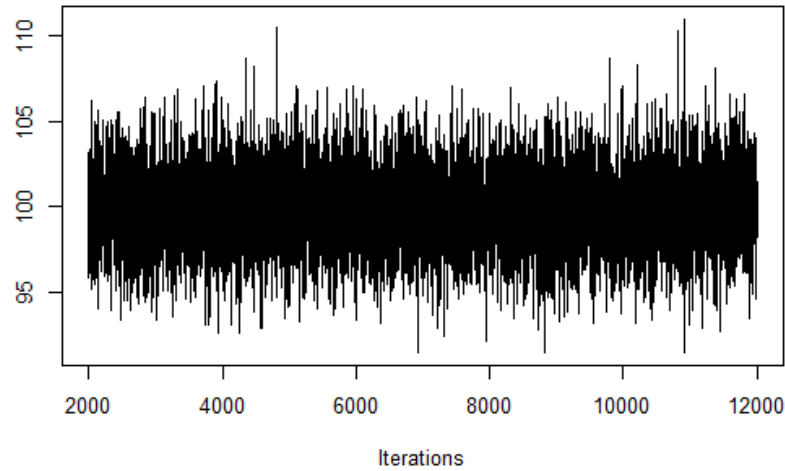
	2.5%	25%	50%	75%	97.5%
beta	95.492	98.466	99.915	101.45	104.45
sigma	7.429	8.848	9.783	10.88	13.67

correlations

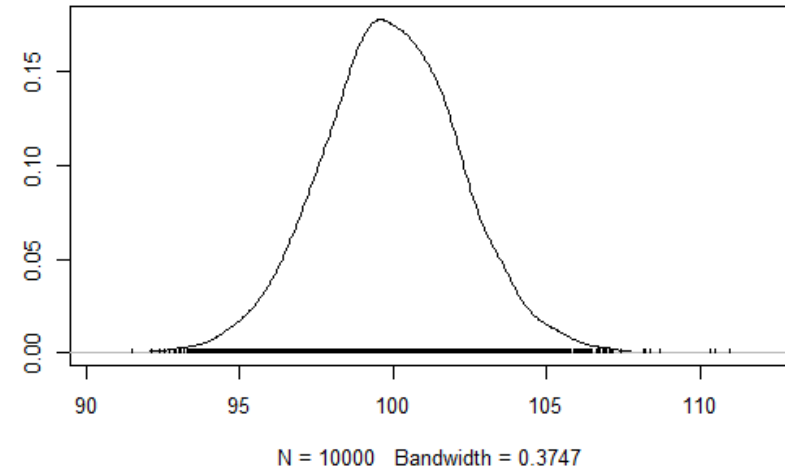
	beta	sigma
beta	1.00000000	-0.02471487
sigma	-0.02471487	1.00000000

# Natural Conjugate Prior

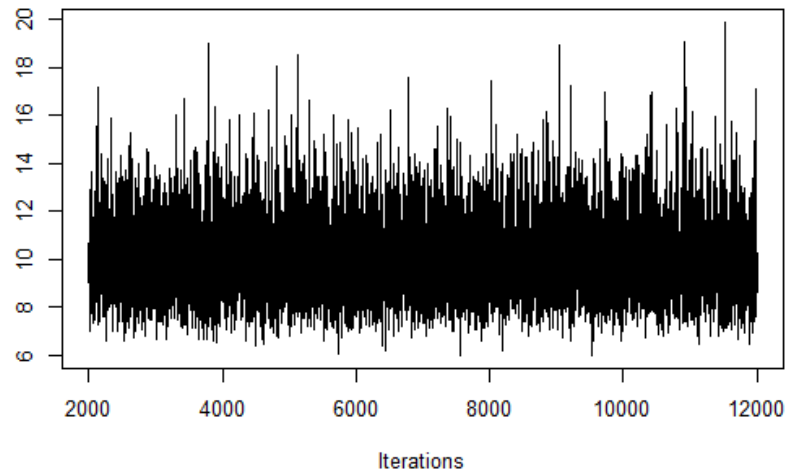
Trace of beta



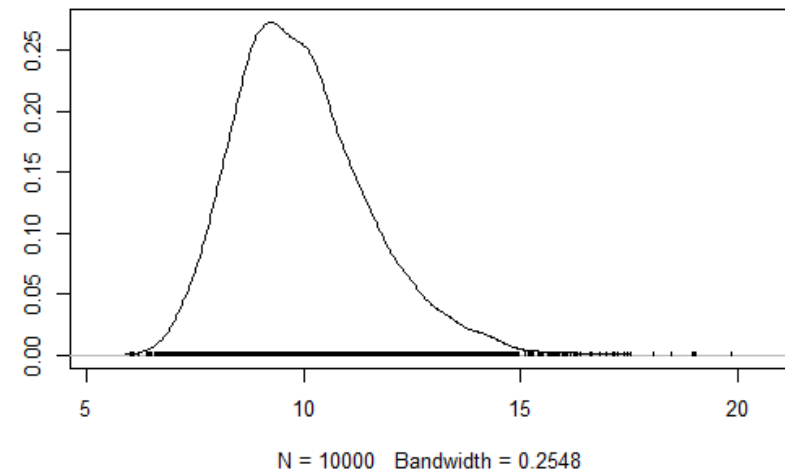
Density of beta



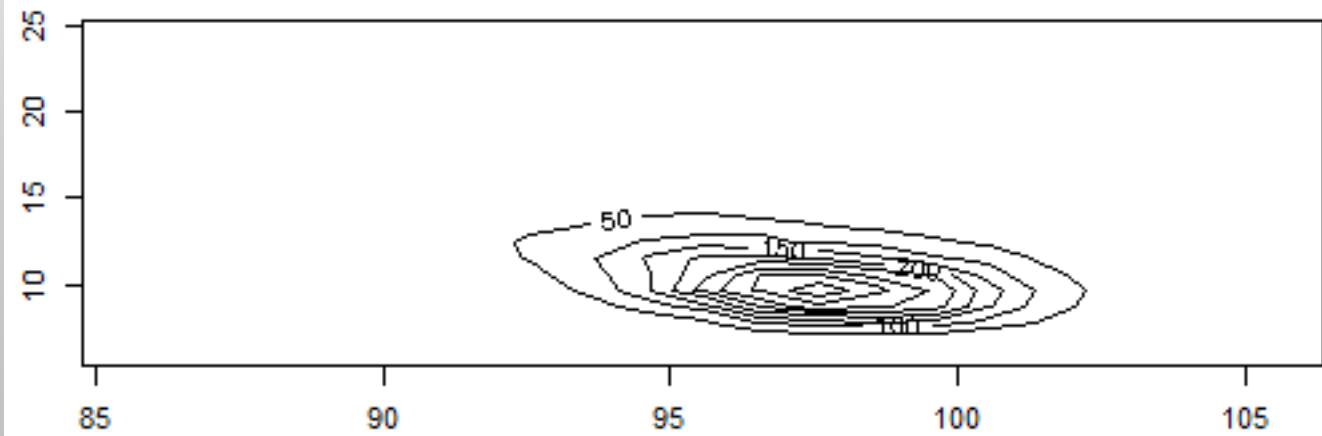
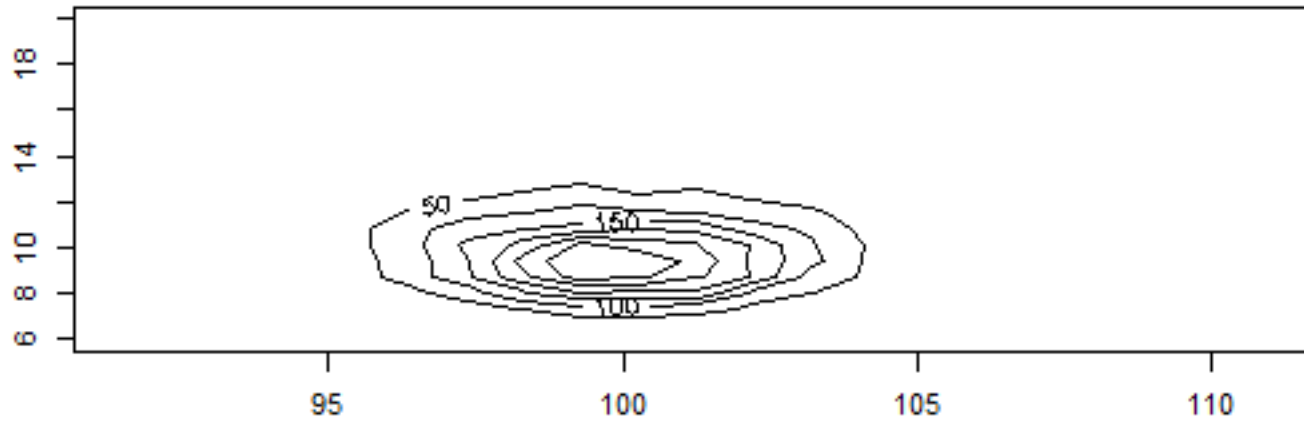
Trace of sigma



Density of sigma



# Natural Conjugate Prior



# Non Informative Prior

```
model
{
  # prior distribution of (beta, sigma)
  # prior sigma distributed as invchi( nu, kappa2 )
  c2 ~ dchisqr( nu )      # chi2( nu )
  t2 <- c2/kappa2        # chi2( nu, 1/kappa2 )
  sigma <- 1/sqrt(t2)    # invchi( nu, kappa2 )

  # prior beta
  beta ~ dnorm( gamma, Phiinv )

  # likelihood
  for( i in 1:nobs ) {
    y[i] ~ dnorm(beta, t2)
  }
}
```

$\text{gamma} = 0$  ,  $\text{Phiinv} = 0.0001$  ,  $\text{nu} = 0.0001$  ,  $\text{kappa2} = 0.0001$

# Non Informative Prior

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
beta	99.91	2.427	0.02427	0.02427
sigma	10.70	1.851	0.01851	0.01955

2. Quantiles for each variable:

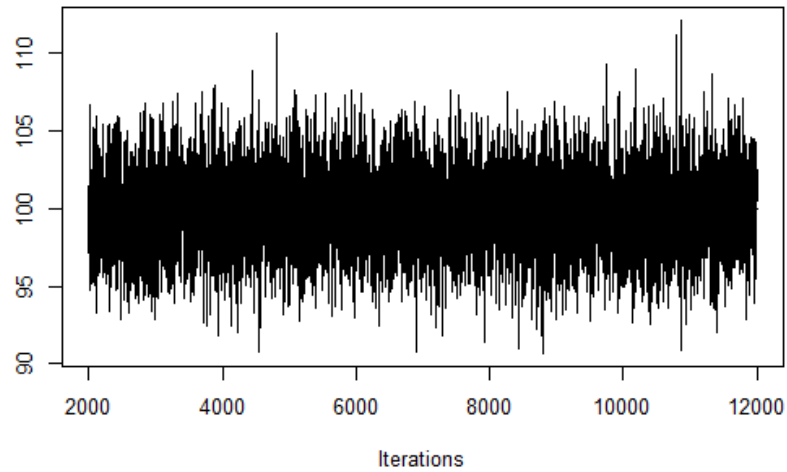
	2.5%	25%	50%	75%	97.5%
beta	95.094	98.331	99.88	101.52	104.73
sigma	7.794	9.384	10.46	11.73	15.02

correlations

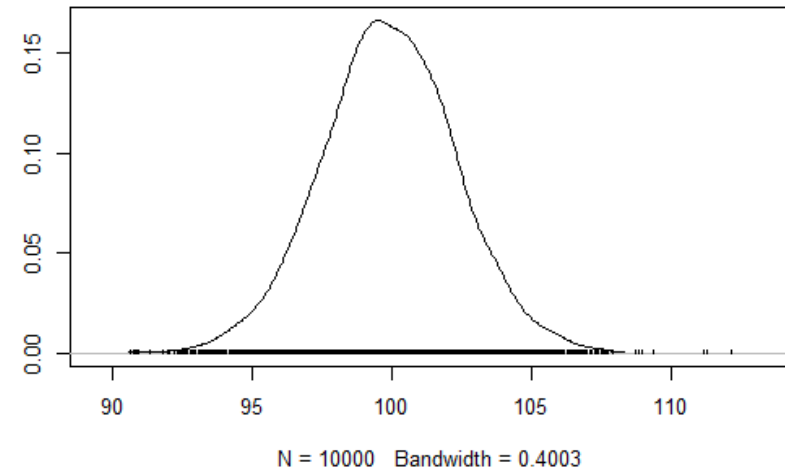
	beta	sigma
beta	1.00000000	-0.03467283
sigma	-0.03467283	1.00000000

# Non Informative Prior

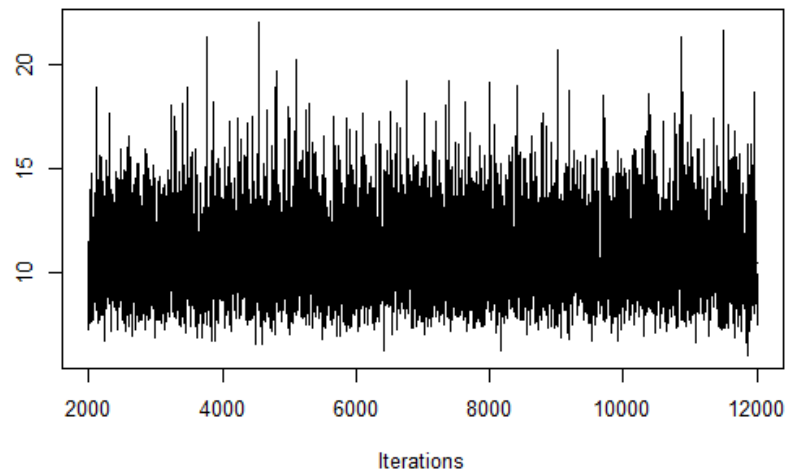
Trace of beta



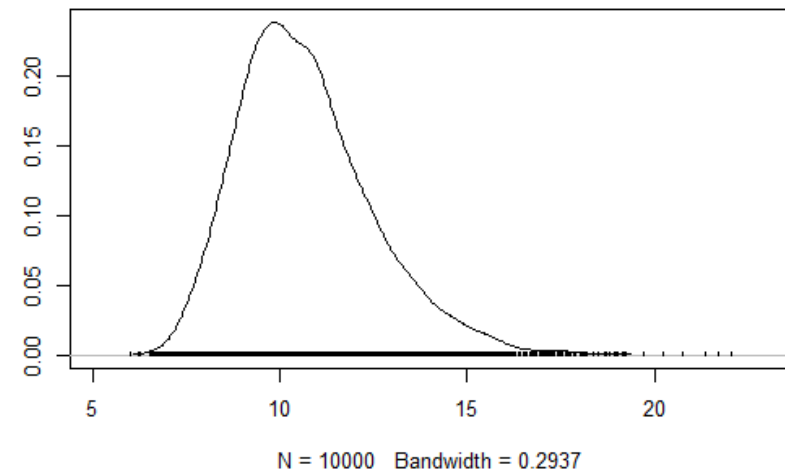
Density of beta



Trace of sigma

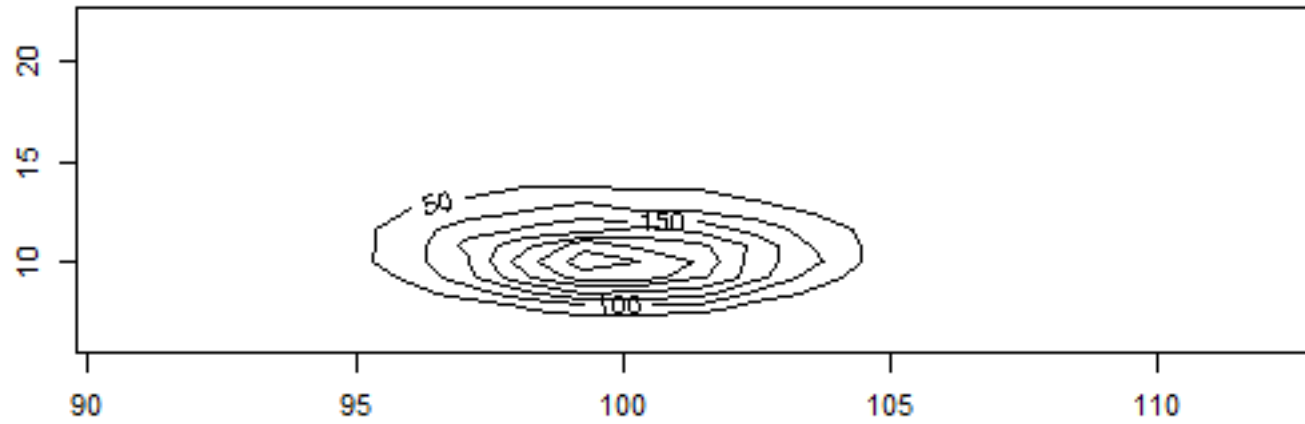


Density of sigma





# Non Informative Prior



# Comparison of Two Means



# free sigma

```
model
{
  # prior distribution of (beta, sigma)
  # prior sigma distributed as invchi( nu, kappa2 )
  c2 ~ dchisqr( nu )      # chi2( nu )
  t2 <- c2/kappa2        # chi2( nu, 1/kappa2 )
  sigma <- 1/sqrt(t2)    # invchi( nu, kappa2 )

  # prior beta
  beta ~ dnorm( gamma, Phiinv )

  # likelihood
  for( i in 1:nobs ){
    y[i] ~ dnorm(beta, t2)
  }
}
```

# free sigma

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
mu1	50.007	0.8146	0.008146	0.008146
mu2	54.994	2.0180	0.020180	0.019431
sigma1	3.610	0.6192	0.006192	0.006478
sigma2	6.806	1.6351	0.016351	0.017658
d12	-4.987	2.1677	0.021677	0.020850

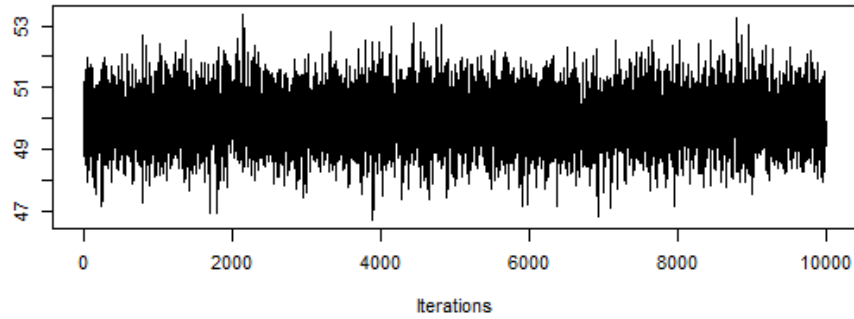
2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu1	48.399	49.472	50.005	50.552	51.6019
mu2	50.987	53.718	55.004	56.276	59.0830
sigma1	2.652	3.174	3.525	3.962	5.0538
sigma2	4.480	5.659	6.523	7.624	10.8522
d12	-9.295	-6.355	-4.996	-3.617	-0.6672

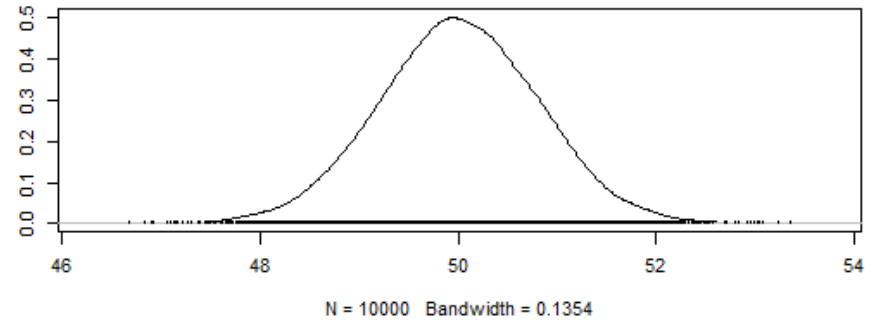
Probability that  $\mu_1 - \mu_2 > 0$  and others  
0.0124

# free sigma

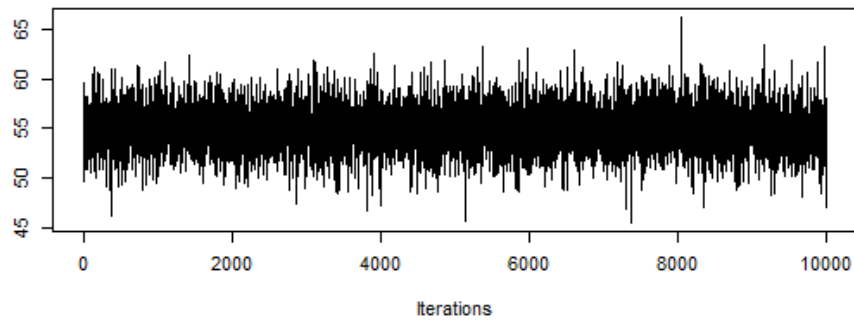
Trace of mu1



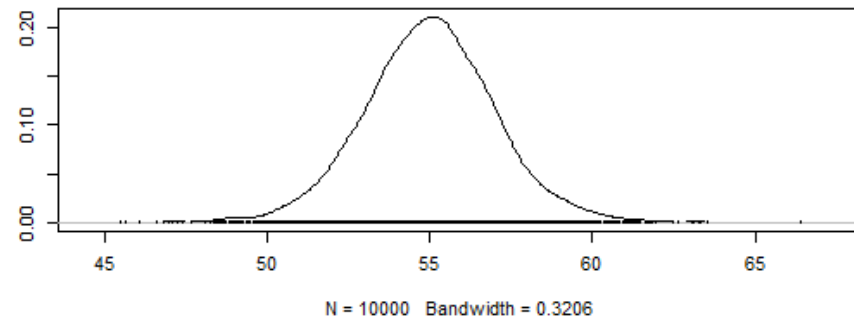
Density of mu1



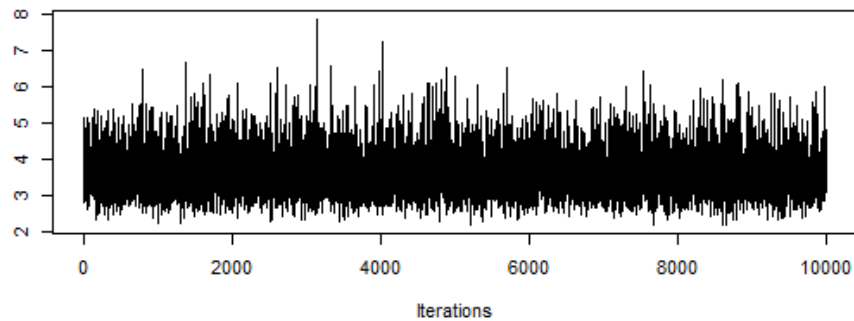
Trace of mu2



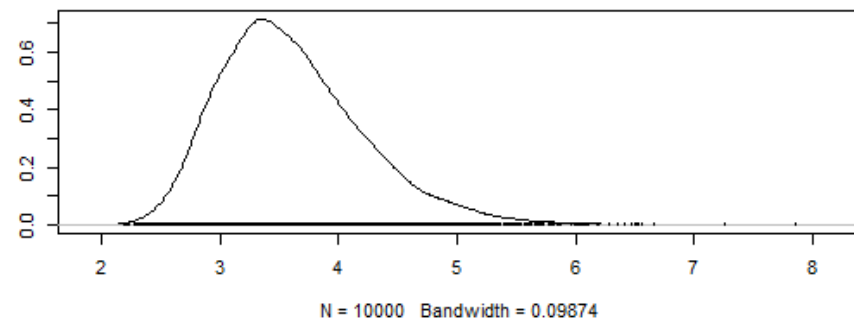
Density of mu2



Trace of sigma1

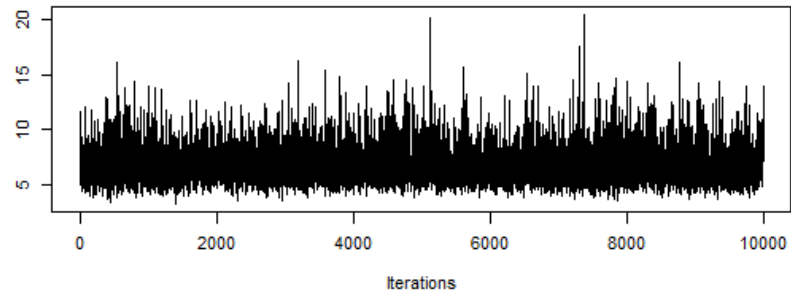


Density of sigma1

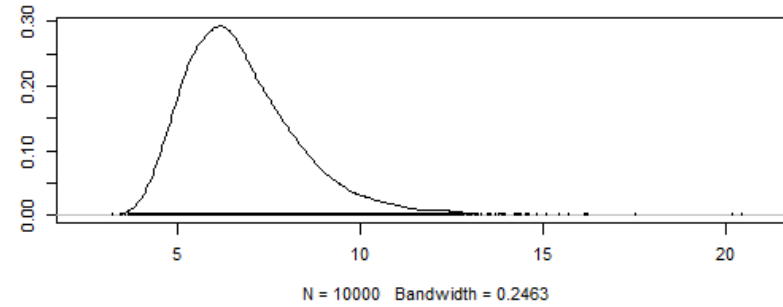


# free sigma

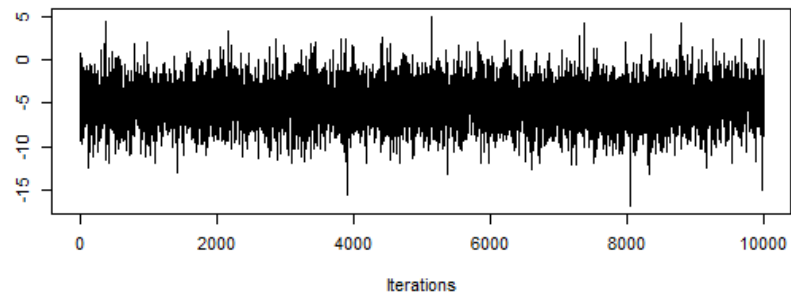
Trace of sigma2



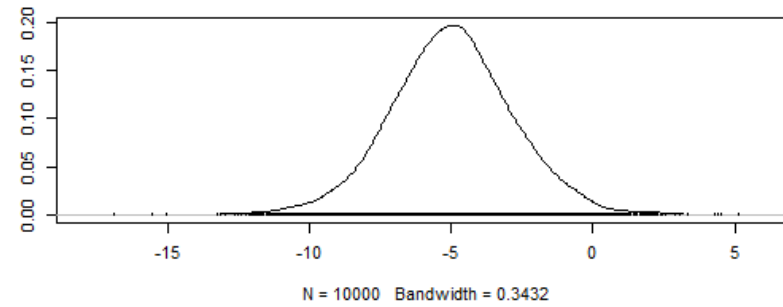
Density of sigma2



Trace of d12



Density of d12



# equal sigma: one way ANOVA

```
model
{
  # prior distribution of (beta, sigma)
  # prior sigma distributed as invchi( nu, kappa2 )
  c2 ~ dchisqr( nu )      # chi2( nu )
  t2 <- c2/kappa2        # chi2( nu, 1/kappa2 )
  sigma <- 1/sqrt(t2)    # invchi( nu, kappa2 )

  # prior beta
  beta ~ dnorm( gamma, Phiinv )

  # likelihood
  for( i in 1:nobs ){
    y[i] ~ dnorm(beta, t2)
  }
}
```

# equal sigma

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
mu1	49.978	1.0897	0.010897	0.010897
mu2	54.977	1.4204	0.014204	0.014204
sigma	4.837	0.6484	0.006484	0.006931
d12	-4.999	1.7791	0.017791	0.017791

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu1	47.834	49.258	49.985	50.698	52.137
mu2	52.166	54.039	54.961	55.921	57.715
sigma	3.747	4.384	4.767	5.231	6.258
d12	-8.540	-6.167	-4.998	-3.846	-1.458

Probability that  $\mu_1 - \mu_2 > 0$  and others  
0.0038



# equal sigma

